AD_____

Award Number: DAMD17-00-1-0417

TITLE: Environmental Exposures at Birth and at Menarche and Risk of Breast Cancer

PRINCIPAL INVESTIGATOR: Jo L. Freudenheim, Ph.D.

CONTRACTING ORGANIZATION: New York State University Research Foundation
Amherst, New York 14228-2567

REPORT DATE: June 2004

TYPE OF REPORT: Final

PREPARED FOR: U.S. Army Medical Research and Materiel Command
Fort Detrick, Maryland 21702-5012

DISTRIBUTION STATEMENT: Approved for Public Release;
Distribution Unlimited

20041123 077

REPORT DOCUMENTATION PAGE

Form Approved
OMB No. 074-0188

Public reporting burden for this collection of information is estimated to average 1 hour per response, including the time for reviewing instructions, searching existing data sources, gathering and maintaining the data needed, and completing and reviewing this collection of information. Send comments regarding this burden estimate or any other aspect of this collection of information, including suggestions for reducing this burden to Washington Headquarters Services, Directorate for Information Operations and Reports, 1215 Jefferson Davis Highway, Suite 1204, Arlington, VA 22202-4302, and to the Office of Management and Budget, Paperwork Reduction Project (0704-0188), Washington, DC 20503

1. AGENCY USE ONLY (Leave blank)	2. REPORT DATE June 2004	3. REPORT TYPE AND DATES COVERED Final (1 Jun 00-31 May 04)	
4. TITLE AND SUBTITLE Environmental Exposures at Birth and at Menarche and Risk of Breast Cancer		**5. FUNDING NUMBERS** DAMD17-00-1-0417	
6. AUTHOR(S) Jo L. Freudenheim, Ph.D.			
7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES) New York State University Research Foundation Amherst, New York 14228-2567 E-Mail: jfreuden@buffalo.edu		**8. PERFORMING ORGANIZATION REPORT NUMBER**	
9. SPONSORING / MONITORING AGENCY NAME(S) AND ADDRESS(ES) U.S. Army Medical Research and Materiel Command Fort Detrick, Maryland 21702-5012		**10. SPONSORING / MONITORING AGENCY REPORT NUMBER**	

11. SUPPLEMENTARY NOTES

12a. DISTRIBUTION / AVAILABILITY STATEMENT Approved for Public Release; Distribution Unlimited	12b. DISTRIBUTION CODE

13. ABSTRACT (Maximum 200 Words)

This population-based study examines early life exposure to environmental pollutants from industrial sites, toxic waste sites and heavily trafficked roadways as risk factors for breast cancer; with a focus on exposure to benzene and phenylalanine hydroxylase (PAHs). We have geocoded 15,340 individual addresses for 3,286 participants in Erie and Niagara counties in New York State. A validation study assessed the positional accuracy of addresses geocoded on the Dynamap2000 using a global positioning system receiver. Overall, geocoding was accurate. Analyses have been completed examining residential proximity to industrial sites contracting with the US Atomic Energy Commission (USAEC), for exposure to total suspended particulates (TSP), and exposure to environmental tobacco smoke (ETS) and breast cancer risk. Proximity to sites contracted by USAEC was not associated with risk. Exposure to TSP in early life was associated with a 2.75-fold increase in risk for postmenopausal women only. There was little evidence of an association between early life exposure to ETS and breast cancer. Clustering analyses identified geographic patterns of residence for breast cancer cases and controls at critical time periods in early life. These results provide evidence that environmental exposures in early life may be important for breast cancer risk.

14. SUBJECT TERMS Cancer etiology, carcinogenesis, environment, geography, molecular epidemiology, gene-environment interaction, early childhood exposures	15. NUMBER OF PAGES 208
	16. PRICE CODE

17. SECURITY CLASSIFICATION OF REPORT Unclassified	18. SECURITY CLASSIFICATION OF THIS PAGE Unclassified	19. SECURITY CLASSIFICATION OF ABSTRACT Unclassified	20. LIMITATION OF ABSTRACT Unlimited

NSN 7540-01-280-5500

Standard Form 298 (Rev. 2-89)
Prescribed by ANSI Std. Z39-18
298-102

Table of Contents

INTRODUCTION

In this population-based study we examined environmental exposures experienced at birth and at menarche as risk factors for breast cancer. We examined location of residence during these potentially sensitive time periods in relation to proximity to industrial sites, toxic waste sites, and heavily trafficked roadways as risk factors for subsequent disease. Residential histories were obtained from all participants in our previously conducted case-control study, which included women age 35-79 with incident, primary, histologically confirmed breast cancer and living in Erie or Niagara counties. Randomly selected controls were frequency matched to cases on age, race, and county of residence. Residence at the time of birth and menarche and potential exposure sites were geocoded into GIS. Primary objectives of this study are: 1) To investigate distance from steel mills, chemical factories, toxic waste sites, other industrial sites and major roadways of the residence of cases and controls at the time of birth and at menarche as risk factors for pre and postmenopausal breast cancer. 2) To examine estimated exposure to benzene and to PAHs as risk factors for pre and postmenopausal breast cancer. 3) To evaluate genetic susceptibility in relation to these exposures and breast cancer risk by examining genetic variability in metabolism by NQO1, GST M1, GST Pi, and CYP1A1. We assessed potential confounding factors including: age, education, income, family history of breast cancer, Quetelet index, body fat distribution, having been breastfed, age at menarche, age at menopause, pregnancy history, lactation and contraceptive history, menstrual cycle length, birth weight, smoking and passive smoke exposure history, and diet and occupational history. In addition to our original aims we also examined 1) clustering of cases compared to controls for residence at birth and at menarche; 2) risk of breast cancer associated with measured total suspended particulates in the air for birth and menarche residence and; 3) risk of breast cancer associated with passive smoke exposure in childhood (another source of PAH exposure). Results of this study are discussed in the text of this report.

BODY OF REPORT

Task 1: Investigate distance from steel mills, chemical factories, gasoline stations, toxic waste sites, and other industrial site of the residence of cases and controls at the time of birth and at menarche as risk factors for pre- and postmenopausal breast cancer.

Task 1 is completed. We have identified and completed data entry for relevant industrial sites and major roadways during time periods under investigation. We have identified additional sources of information regarding historical sources of the exposures of interest and their locations and amounts. We have verified and geocoded residential histories of study participants. We have completed the geocoding for study participants for their residence at the time of their birth, at menarche, when they had a first birth and 10 and 20 years before diagnosis (cases) or interview (controls). In total, we have geocoded approximately 20,000 addresses in Erie and Niagara counties. In addition, we have conducted a validation study of the positional accuracy of geocoded residences. Results of this validation study will be published in the journal Epidemiology in July 2003 (see APPENDIX I). We have also completed data analysis examining early life proximity to industrial sites contracted by the United States Atomic Energy Commission in relation to risk of breast cancer in adult life. Currently a manuscript for these analyses is in preparation and will be submitted for publication in the next year.

Two abstracts from the work on this task were presented at the annual meeting of the Society for Epidemiologic Research in Atlanta, Georgia, June 11-14, 2003 and the abstracts were published in a supplement of the American Journal of Epidemiology. They are: "Residential Proximity to Chemical or Primary Metal Industry and the Risk of Breast Cancer in Western New York," and "Clustering of Lifetime Residence and Breast Cancer Risk in Western New York." Copies of the abstracts are in APPENDIX V.

We completed a GIS-based spatial and temporal analyses for residences of breast cancer cases and controls at early life. Since we found strong evidence of spatial clustering for cases at early time periods, we will continue with the next step which is to estimate breast cancer risk associated with environmental exposures at those early residences. Epidemiologic investigations on the evaluation of environmental risk factors and the estimation of breast cancer risk associated with lifetime residential history will be performed with the aid of GIS and spatial perspectives. A manuscript is in preparation regarding the clustering of lifetime residence and breast cancer risk using exploratory spatial analysis tools based on these lifetime residential history data. Descriptions of findings follow.

A. Analysis of Geographic Clustering of Cases and Controls by Period of the Lifetime

First, we completed a GIS-based spatial and temporal analysis for residences of breast cancer cases and controls at early life and found strong evidence of spatial clustering for cases during this time. A paper, "Geographic Clustering of Residence in

Early Life and Subsequent Risk of Breast Cancer," has been accepted for publication in Cancer Causes and Control. A copy of the complete manuscript is included in APPENDIX II.

Second, we examined breast cancer risk associated with lifetime residential history to identify spatio-temporal patterns of risk surfaces in a population-based case-control study of breast cancer. With a growing interest in early or lifetime exposures to breast cancer risk, a life-course approach was adapted to see whether environments in early life or biological processes around critical events in a life-course may be related to disease in adulthood (Kuh and Ben-Shlomo, 1997; Barker, 1992). We explored the use of density estimation methods in epidemiologic studies as GIS-based exploratory spatial analyses, and obtained risk surfaces using several measures such as smoothed ratio and standardized difference. These risk surfaces were produced and compared between residences for pre-menopausal and post-menopausal women. It provides risk surfaces from lifetime residential history, thus more accurate estimates than density surfaces based on only current residential location.

We used six temporal groups, place of birth, the primary residence during the period of menarche, and the primary residence during women's first birth, residence 10 and 20 years prior to diagnosis for the cases and prior to interview for the controls, and current addresses. Geocoding of residential locations in six temporal groups are essential parts of this study which enables us to record each individual's locational information as x and y coordinates to be used in further spatial analyses. Overall address matching rates were 92.5%. Table 1 is a summary table showing the final numbers of cases and controls for those six events by menopausal status.

We first identified areas with higher than average densities of breast cancer cases in the study area based on the relative densities of cases to controls. Figure 1 shows the residential locations of breast cancer cases and controls in western New York. There are 4,808 residential locations for cases with 1334 pre-menopausal and 3470 post-menopausal residences, while there are 8,580 residential locations for controls with 2559 pre-menopausal and 6010 post-menopausal residences. We produced two maps of risk surface based on residential locations of cases and controls in Figure 1. Figure 2 shows risk surfaces of pre- and post-menopausal residences, and depict only areas of high case densities in the study area (ratio greater than 0.5). For instance, areas with ratio greater than 0.76 (with contours) indicate two times of increased breast cancer risk.

Second, standardized difference in case and control density is obtained to assess variability of risk surfaces. Figure 3a and 3b depict areas of greater than two standard deviations (with contours), and those statistically significant areas with images (over the critical value 3.5). There are several areas of interest for pre-menopausal residences; two in the center of the study area, and one in rural area, while only one in rural area was detected for post-menopausal residences. These are statistically significant areas exceeding the critical value at $\alpha=.01$, indicating density of cases are significantly higher than that of controls. However, interpretation should be cautious on the ones appeared in rural area for both pre-menopausal and post-menopausal residences. As seen in Figure 1,

6

there are very sparse residences in those areas. Although it is suggested that the difference between case and control density is significantly different, it is not reliable because small sample size may influence on the results.

Finally, we were interested in finding time periods contributing to this result. Standardized difference is obtained for each temporal group, and pre- and post-menopausal residences separately. Figure 4 and 5 shows risk surfaces difference in space and time. As seen in Figures 4 and 5, areas greater than 2 S.D. are illustrated for each temporal group with darker images as statistically significant areas of high case density. Testing for significance are performed and attached as p-values. We found three time periods of interests; birth, menarche and 10 years for pre-menopausal residences. These areas are significantly different at α.=0.1 and indicate areas of high case densities. For post-menopausal residences, we found one significant time period (20 years) at the .1 level. Once again ones in rural area for 10 years in Figure 4 and 20 years in Figure 5 seem to appear due to small sample size of the area. A manuscript describing these results is in preparation.

Table 1. Residential history of breast cancer cases and controls

	A-Complete Addresses						B-Missing		C-Total
	Cases			Controls					(%A/C)
	Pre-	Post-	All	Pre-	Post-	All	Cases	Controls	
Birth	160	283	505	345	521	804	127	189	1642(80)
Menarche	204	386	673	469	757	1143	98	154	2068(88)
First birth	181	371	616	435	782	1153	97	167	2033(87)
20 years*	210	672	882	413	1201	1614	96	157	2749(91)
10 years*	258	717	975	501	1266	1767	74	133	2949(93)
Current	327	826	1157	619	1469	2099	12	18	3286(99)
Total	1334	3470	4808	2559	6010	8580	504	818	14727(91)

* 20 and 10 years prior to diagnosis or control selection respectively
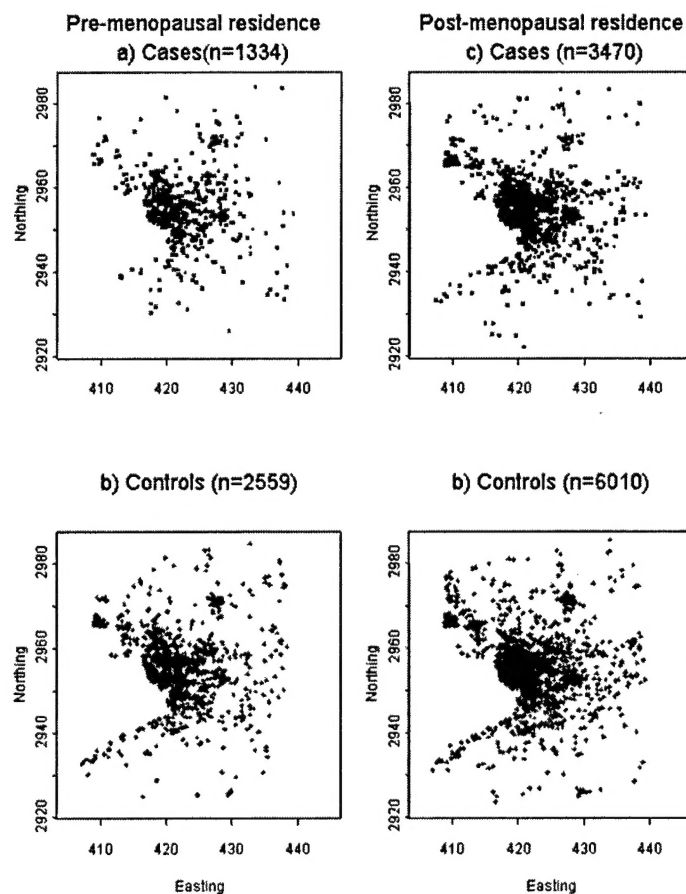
Figure 1. Geographic distribution of breast cancer cases in western New York: All residential locations of breast cancer cases and controls included in the analysis
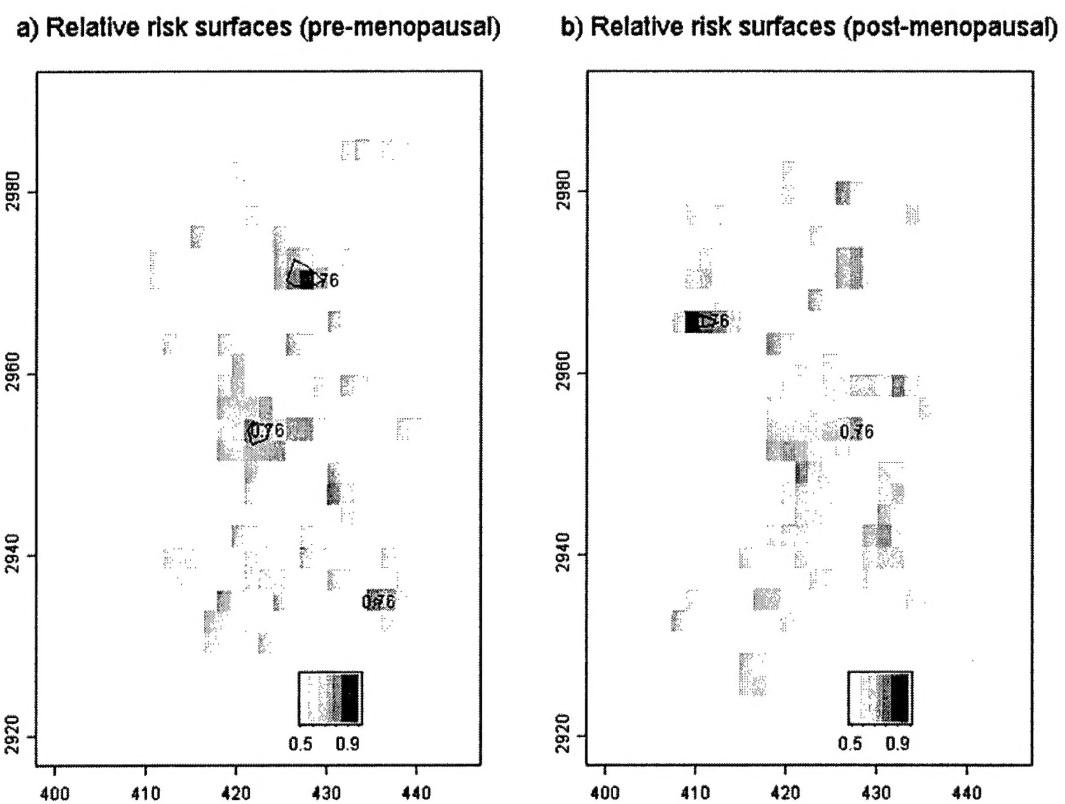
Figure 2. Risk surfaces for pre-menopausal and post-menopausal residence

a) Difference surfaces (pre-menopausal)   **b) Difference surfaces (post-menopausal)**

Figure 3. Difference in risk surfaces for pre-menopausal and post-menopausal residence

Figure 4. Risk surface difference in space and time: pre-menopausal residence

Figure 5. Risk surface difference in space and time: post-menopausal residence

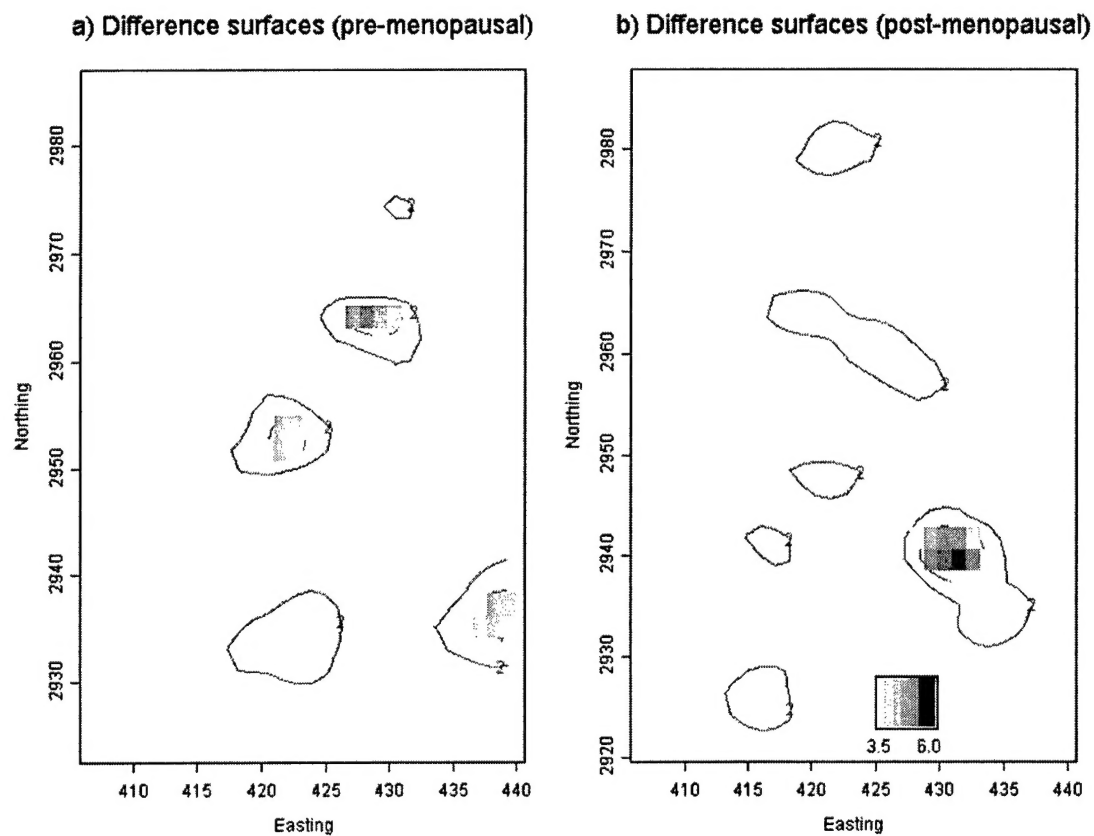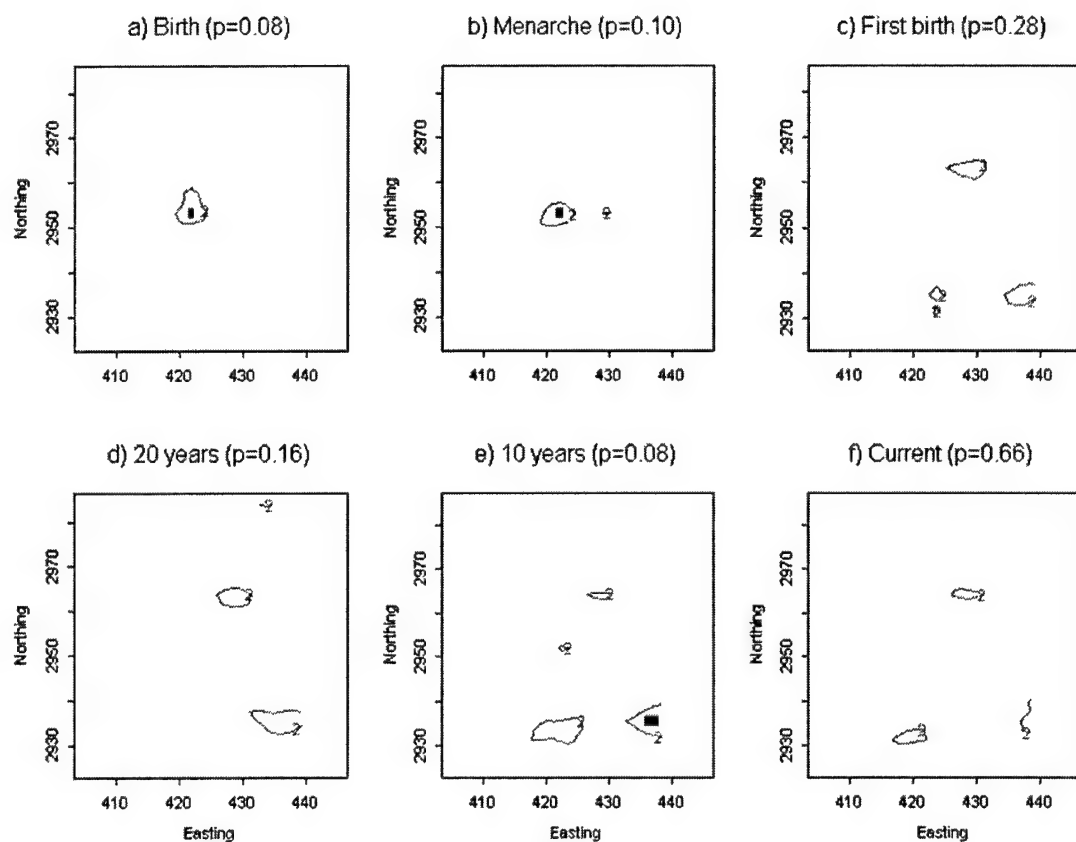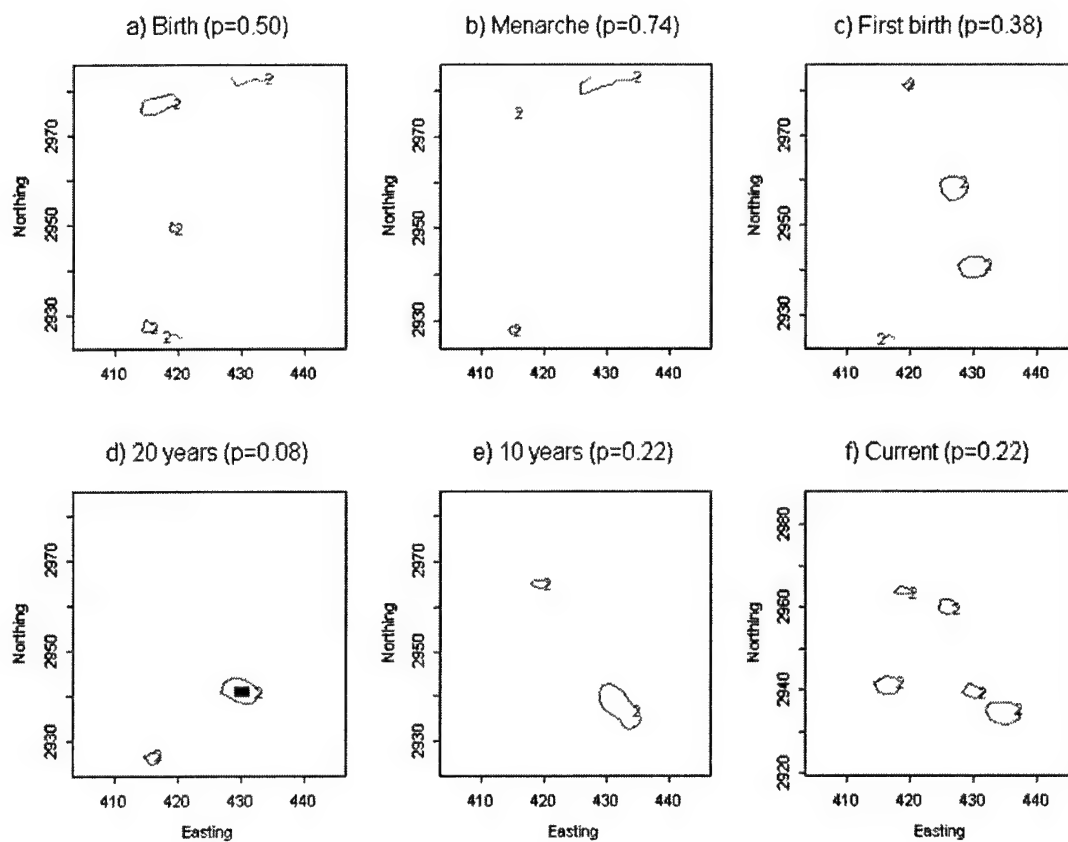## B. Examination of Breast Cancer Risk in Relation to Residential Proximity to Industrial Sites Contracted by the U.S. Atomic Energy Commission

Ionizing radiation is a well recognized human mammary carcinogen. Numerous studies including those of Japanese atomic bomb survivors and tuberculosis patients treated with radiation have shown that breast epithelium is radiosensitive to external radiation. In addition, exposure at early age appears to be particularly important. Less is known about the effects of internal radiation on breast epithelium. The epidemiologic evidence regarding internal emitters and breast cancer comes primarily from radium dial workers and from German patients treated with high doses of radium-224 for ankylosing spondylitis and tuberculosis. In the German patients, adult women treated with radiun-224 had an SIR of 1.77 for breast cancer, while women < 21 years of age when treated with radium-224 had an SIR of 9.4, further suggesting that early life exposures may be important in mammary carcinogenesis, albeit at high doses. The effects of low-dose exposure to the general population, however, have not been demonstrated and are generally extrapolated from high-dose exposures.

The general population is exposed to low-dose internal and external radiation from both natural and anthropogenic sources. Natural sources include decay of uranium present in soil, dissolved in water, and absorbed by plants and animals that are consumed. Man-made sources include therapeutic radioactive isotopes, consumer products such as tobacco and smoke detectors, the nuclear power industry, and the military nuclear industry.

In the 1940's and 1950's, the United States Atomic Energy Commission (USAEC) and its predecessor the Manhattan Engineering Project (currently, the United States Department of Energy) contracted with numerous private industries to process uranium for the burgeoning nuclear program. In Erie and Niagara Counties in Western New York State, 13 such industrial sites contracted with the USAEC to enrich uranium, mill uranium metal, or to store radioactive waste. In addition to these USAEC activities, these sites were also engaged in commercial industrial activities such as steel and chemical manufacturing. In this case-control study we examined women's residential proximity to these sites at the time of their birth, at menarche, and when they had their first birth in relation to breast cancer in adult life. These industrial sites were examined because they were relatively close to residential neighborhoods and were engaged in processing radioactive material that resulted in residual environmental contamination at most sites. We have focused on early life exposure because it appears that it is the critical time period in breast development when breast epithelium is particularly sensitive to effects of ionizing radiation.

For this study, we postulated that women exposed in early life to radiation from uranium-238, uranium-235, radium-226 and thorium-232 from USAEC site activities would be more likely to develop breast cancer than women without these early life exposures. Specifically, we hypothesized that women born in close proximity to USAEC sites would be more likely to develop breast cancer than women born further away. In addition, we also predicted that women who resided in close proximity at the time of menarche and at first birth would also have increased odds of breast cancer compared to women residing further away from these sites at menarche and first birth.

Methods

A population-based, case-control study was conducted to evaluate the proposed hypotheses. Cases consisted of 1,166 women aged 35-79 living in Erie or Niagara County diagnosed with histologically confirmed, primary, incident breast cancer between the years 1996 and 2001. Controls under 65 years of age were randomly selected from the New York State Department of Motor Vehicles driver's license list and controls 65 and over were randomly selected from the Healthcare Financing Administration Medicare rolls. Controls ($n = 2,105$) were frequency matched to cases on age, race, and county of residence. The response rates for the cases were 59% and 35% for the controls. Refusal to participate was the he most common reason for both cases and controls. These estimates of response are somewhat conservative in that they include in the denominator 18% of cases and 45% of controls where eligibility could not be determined. With these individuals removed from the denominator, the response rates were 72% for cases and 79% for controls. The true response rate, however, most likely lie somewhere between these two estimates for cases and controls.

Extensive in-person interviews and self-administered questionnaires were used to ascertain medical history, diet, alcohol consumption, smoking history (including passive smoke exposure), residential history, and occupational history. Each participant listed all their residencies for their lifetime starting with the address at the time of interview. When a subject could not provide a complete address in Erie or Niagara County, Polk directory and city directory were searched to find this missing information. These histories were used to locate each subject's residence at birth, menarche, and first birth.

For the proximity at birth analyses, cases and controls were restricted to those with birth addresses in Erie or Niagara Counties. Of these, a further 241 cases and 380 controls were excluded from these analyses because they were born prior to the period of USAEC activities in this region (1942-1956). A total of 261 cases and 424 controls were included in the birth analyses.

For the analyses assessing exposure at menarche, cases and controls were restricted to those with an address in Erie or Niagara Counties at the time of menarche during the period of USAEC activities, leaving 581 cases and 918 controls. For the analyses assessing proximity at first birth, cases and controls were restricted to those with an address in Erie or Niagara Counties at the time of birth of their first child. Of these, one case and 14 controls were excluded because their first birth occurred prior to USAEC site activities for a total of 615 cases and 1,139 controls.

Exposure Assessment

Proximity to USAEC industrial sites was used as a surrogate for exposure to radioactive pollution emanating from these sites. Proximity to USAEC sites was calculated in a two step process. All birth, menarche, and first birth addresses were geocoded with ArcView 3.2 (ESRI, Inc., Redlands, CA) on the Dynamap2000 reference theme (Geographic Data Technologies, Inc., Lebanon, NH). The addresses of all 13 USAEC sites were also geocoded onto the same reference theme. An extension to ArcView 3.2 was used to calculate a distance matrix for each address to each of the 13 USAEC sites.

Subjects could theoretically be exposed to pollutants from all 13 sites; however, we used the closest site to calculate proximity with the rationale that the closest site would

14

contribute the majority of exposure. An algorithm was created in SAS (SAS Institute, Inc., Cary, NC) to determine the closest site dependent on the year that the subject's critical event occurred (i.e., birth, menarche, or first birth) and the years in which the USAEC sites were actively engaged in production or processing of radioactive material. This was done to ensure that only active plants were used to determine proximity. For example, there were only two plants operating in 1942. Consequently, for participants born in 1942, their closest site would be one of those two plants. We opted not to use those subjects born prior to USAEC activities as the truly unexposed because of the potential to introduce a birth cohort effect. Furthermore, there were only two premenopausal women who were born prior to 1942, precluding a comparison of this type for the premenopausal women.

Radiological surveys for 9 of the 13 sites were obtained from the United States Army Corps of Engineers. Radiological surveys were conducted primarily by the National Laboratory at Oak Ridge to assess the amount of radiological contamination present on these sites for the Formerly Utilized Sites Remedial Action Program. These surveys were used to estimate the potential for off-site contamination and to estimate radiologic dose from that contamination. Whole body dose was estimated with the RESRAD 7.0 code (Argonne National Laboratory, Argonne, IL) using the default values and the average soil concentrations of uranium-238, uranium-235, radium-226, and thorium-232 at the Linde Ceramic Plant, which had the highest ground concentrations of uranium-238, uranium-235, radium-226, and thorium-232 of all the sites. Average soil concentrations of these radionuclides were derived from the radiological surveys. Radiological surveys for the remaining four sites either were never done or could not be located.

Statistical Analyses

The Student's T-test and Pearson's Chi-square test were used to compare demographic and reproductive characteristics between cases and controls. The distance to the closest site was categorized into quartiles based on the distribution of distance in the controls. The closest quartile was further divided in two at the midpoint to provide higher resolution for the closest distances. Logistic regression was used to calculate odds ratios (OR) and 95% confidence intervals (95% CI) for each quartile compared to the furthest quartile. Multiple logistic regression was used to assess potential confounding by age at interview, race, education, age at menarche, parity, age at first birth, previous benign breast disease, family history of breast cancer, body mass index (height (m)/weight $(kg)^2$), and age at menopause for postmenopausal women. All models were stratified by menopausal status to assess effect measure modification. P for trend statistics were determined by the p-value for the coefficient of the continuous exposure variable, while adjusting for covariates.

Results

Descriptive characteristics of subjects stratified by menopausal status are depicted in Table 1. Cases were born on average 1 km closer to a USAEC site than controls. For postmenopausal women, cases were born about 0.5km closer than controls.

The associations between proximity to USAEC sites at birth and subsequent breast cancer are shown in Table 2. In premenopausal women, proximity within 3.3 kilometers of a USAEC site was suggestive of a slight increase in the odds ratio of breast cancer

15

(adjusted OR = 1.69, 95% CI = 0.68-4.21), although there were few women in this category. There was no evidence of a linear association with distance. A similar pattern was also seen with the postmenopausal women with a slightly raised breast cancer OR for subjects with birth residences within 3.3 km of their closest USAEC site (adjusted OR = 1.30, 95% CI = 0.43-3.99). Nevertheless, the confidence intervals for both pre- and postmenopausal women are also consistent with no increase in breast cancer risk.

As previously mentioned, the 13 USAEC sites were engaged in various uranium processing activities and the potential for the general population to be exposed to radionuclides and radiation from these sites may have differed depending on the activities at that site. However, proximity to either waste storage facilities or uranium enrichment/metal processing sites at the time of birth was not associated with breast cancer in either pre- or postmenopausal women (data not shown).

In Table 3, ORs and 95% CI for proximity of subjects' addresses at menarche to the closest USAEC site are shown. There were no consistent associations with proximity to these sites and risk for this time period of exposure. For premenopausal women, residing within 3.3 km of a USAEC site at menarche compared with women residing 15 km or greater at menarche the OR was 1.42 (95% CI = 0.46-4.34). There was, however, an apparent reduction in the OR for distances between 3.3 and 10.2 km. For postmenopausal women, proximity <15 km was associated with a reduction in the OR. For postmenopausal women residing within 3.3 km of an USAEC sites, the OR was 0.54 (95% CI = 0.28-1.02). Proximity of residence at first birth was also not consistently associated with subsequent breast cancer in either pre- or postmenopausal women (Table 4).

We used the radiological surveys and RESRAD 7.0 code (Argonne National Laboratory, Argonne, IL.) to estimate whole body doses of ionizing radiation from on-site contamination with uranium-238, uranium-235, radium-226, and thorium-232. In a worst case scenario, an individual residing on a premises of a $1000m^2$, with the average concentration of radionuclides estimated from the Linde Plant, would have an Effective Dose Equivalent of 0.42 mSv/year, which is within the range of background radiation exposure experienced by the general population (3.6 mSv/yr).

Table 1. Descriptive Characteristics (means (SD) and percentages) of Study Participants Born in Erie and Niagara Counties between 1942 and 1964.

Variable	Premenopausal			Postmenopausal		
	Cases (n=160)	Controls (n=281)	P-value	Cases (n=104)	Controls (n=143)	P-value
Age (years)*	44.26 (4.51)	43.68 (4.37)	0.18	53.07 (3.05)	50.39 (3.29)	<0.0001
Education (years)*	13.71 (2.15)	14.24 (2.30)	0.01	14.07 (2.39)	13.64 (2.31)	0.16
Age at Menarche (years)*	12.52 (1.48)	12.56 (1.65)	0.79	12.08 (1.58)	12.45 (1.61)	0.07
Age at Menopause (years)*	--	--		48.32 (4.29)	45.41 (5.67)	<0.0001
Age at First Birth (years)*	19.76 (10.03)	21.58 (10.74)	0.80	19.14 (9.92)	19.78 (9.72)	0.61
Body Mass Index*	27.36 (7.15)	27.23 (6.19)	0.84	25.59 (5.43)	28.89 (7.32)	0.71
Distance to Closest Site (km)*	11.30 (6.21)	12.58 (7.96)	0.06	10.19 (5.89)	10.76 (7.39)	0.50
Parity (%)						
0 births	15%	17%		18%	17%	
1-2 births	59%	54%		51%	46%	
3+ births	26%	28%	0.62	31%	37%	0.59

Table 1, continued. Descriptive Characteristics (means (SD) and percentages) of Study Participants Born in Erie and

Niagara Counties between 1942 and 1964.

Variable	Premenopausal			Postmenopausal		
	Cases (*n*=160)	Controls (*n*=281)	P-value	Cases (*n*=104)	Controls (*n*=143)	P-value
First-degree Relative with	23%	9%	0.0002	16%	7%	0.02
Breast Cancer (% yes)						
Benign Breast Disease (%	34%	22%	0.01	39%	25%	0.02
yes)						

* Mean (standard deviation)

** T-tests were used to calculate p-values for continuous variables and χ^2 tests were used for categorical variables

18

Table 2. Odds Ratios and 95% Confidence Intervals for Birth Address Distance to the Closest US Atomic Energy Commission Site.

Distance (km)	Premenopausal				Postmenopausal			
	Cases (n=160)	Controls (n=281)	Crude OR (95%CI)	Adjusted OR (95%CI)*	Cases (n=104)	Controls (n=143)	Crude OR (95%CI)	Adjusted OR (95%CI)*
>14.87	35	90	1.00	1.00	14	28	1.00	1.00
10.21-14.86	57	63	2.33 (1.37-3.95)	2.20 (1.25-3.89)	29	40	1.45 (0.65-3.23)	1.80 (0.68-4.71)
6.60-10.20	26	60	1.11 (0.61-2.04)	0.96 (0.51-1.82)	35	33	2.12 (0.95-4.71)	1.69 (0.66-4.32)
3.30-6.60	30	54	1.43 (0.79-2.59)	1.36 (0.73-2.55)	15	21	1.43 (0.57-3.59)	1.56 (0.51-4.83)
<3.30	12	14	2.20 (0.93-5.23)	1.69 (0.68-4.21)	11	21	1.05 (0.40-2.77)	1.30 (0.43-3.99)
P for trend				0.2236				0.6613

* Adjusted for age, race, education, previous benign breast disease, parity, age at first birth, age at menarche, BMI, age at first birth, family history of breast cancer, and age at menopause for postmenopausal women only.

Table 3. Odds Ratios and 95% Confidence Intervals for Menarche Address Distance to the Closest US Atomic Energy Commission Site.

Distance (km)	Premenopausal				Postmenopausal			
	Cases (n=204)	Controls (n=386)	Crude OR (95%CI)	Adjusted OR (95%CI)*	Cases (n=377)	Controls (n=532)	Crude OR (95%CI)	Adjusted OR (95%CI)*
>14.87	78	147	1.00	1.00	95	122	1.00	1.00
10.21-14.86	62	88	1.32 (0.87-2.03)	1.30 (0.82-2.04)	98	149	0.85 (0.58-1.22)	0.72 (0.48-1.06)
6.60-10.20	33	96	0.65 (0.40-1.05)	0.65 (0.39-1.08)	90	130	0.89 (0.61-1.30)	0.70 (0.47-1.05)
3.30-6.60	25	46	1.02 (0.59-1.79)	0.94 (0.52-1.70)	72	96	0.96 (0.64-1.45)	0.76 (0.49-1.17)
<3.30	6	9	1.26 (0.43-3.67)	1.42 (0.46-4.34)	22	35	0.81 (0.44-1.47)	0.54 (0.28-1.02)
P for trend				0.3978				0.1937

* Adjusted for age, race, education, previous benign breast disease, parity, age at first birth, age at menarche, BMI, age at first birth, family history of breast cancer and age at menopause for postmenopausal women only.

20

Table 4. Odds Ratios and 95% Confidence Intervals for First Birth Address Distance to the Closest US Atomic Energy Commission Site.

Distance	Premenopausal					Postmenopausal				
	Cases	Controls	Crude OR	Adjusted OR		Cases	Controls	Crude OR	Adjusted OR	
(km)	$(n=181)$	$(n=371)$	(95%CI)	(95%CI)*		$(n=434)$	$(n=768)$	(95%CI)	(95%CI)*	
>14.87	79	171	1.00	1.00		133	211	1.00	1.00	
10.21-14.86	36	77	1.01 (0.63-163)	0.95 (0.56-1.59)		109	247	0.70 (0.51-0.96)	0.70 (0.50-0.97)	
6.60-10.20	38	78	1.06 (0.66-1.69)	1.14 (0.70-1.86)		116	174	1.06 (0.77-1.46)	1.01 (0.72-1.41)	
3.30-6.60	25	39	1.33 (0.79-2.45)	1.34 (0.74-2.43)		57	104	0.87 (0.59-1.28)	0.86 (0.57-1.30)	
<3.30	3	6	1.08 (0.26-4.44)	0.89 (0.20-4.06)		19	32	0.94 (0.51-1.73)	1.09 (0.57-2.08)	
P for trend				0.7820					0.2726	

*Adjusted for age, race, education, previous benign breast disease, parity, age at first birth, age at menarche, BMI, age at first birth, family history of breast cancer and age at menopause for postmenopausal women only.

21

C. Proximity to Chemical or Primary Metal Industrial Sites

Women living in urban environments are at greater risk of breast cancer than those in rural settings; this difference is not well understood. We conducted a study to examine environmental exposures 10 and 20 years prior to diagnosis (cases) or interview (controls) in relation to breast cancer risk, in particular risk associated with: 1) residential proximity to chemical industry sites; 2) residential proximity to primary metal industry sites. It's a population-based case control study. Cases were women, age 35-79 with incident, primary, histologically confirmed breast cancer living in Erie or Niagara counties; controls were population based, frequency matched to cases on age, race and county. Self-reported lifetime residential histories were collected, and missing address information supplemented with Polk Directory searches. 863 cases and 1579 controls with complete residential addresses for the periods 10 and 20 years prior to diagnosis (or interview for controls) were included in these analyses. Industrial directories for New York State for 1978 and 1988, were used to identify chemical and primary metal factories operating in this region. The chemical facility in our study includes Standard Industrial Classification (SIC) groups 28 (chemicals and allied products), 29 (petroleum refining and related industries), and 30 (rubber and miscellaneous plastics products); and primary metal facility includes SIC 33. We used ArcView3.2 (using GDT/Dynamap as the base map) to geocode the addresses. The locations of industrial sites and residences are list in Figure 1 and 2. Quartiles were created to categorize the distance from residential address to the closest industrial site; women living within 0.25 mile of a facility were put in a separate category. We used logistic regression to calculate the odds ratios and 95% confidence intervals, adjusting for age, education, race, BMI, age at first birth, age at menarche, age at menopause (postmenopausal women), parity, first-degree relative with breast cancer, and previous benign breast disease. For both pre and postmenopausal women, no evidence that living close to chemical or primary metal facility 10 and 20 years ago was associated with increased risk (Tables 1 to 4). In this study, we used proximity to estimate exposure. However, the complexity of the chemical mixtures from different sites likely leads to exposure misclassification. While our measure of exposure to any single compound is crude, the real world exposure is generally to mixtures of compounds. In this study we found no effect of exposure to these mixtures in the recent decades on breast cancer risk.

Figure 1. Locations of Industries and Residence by Case-control Status, Premenopausal Women
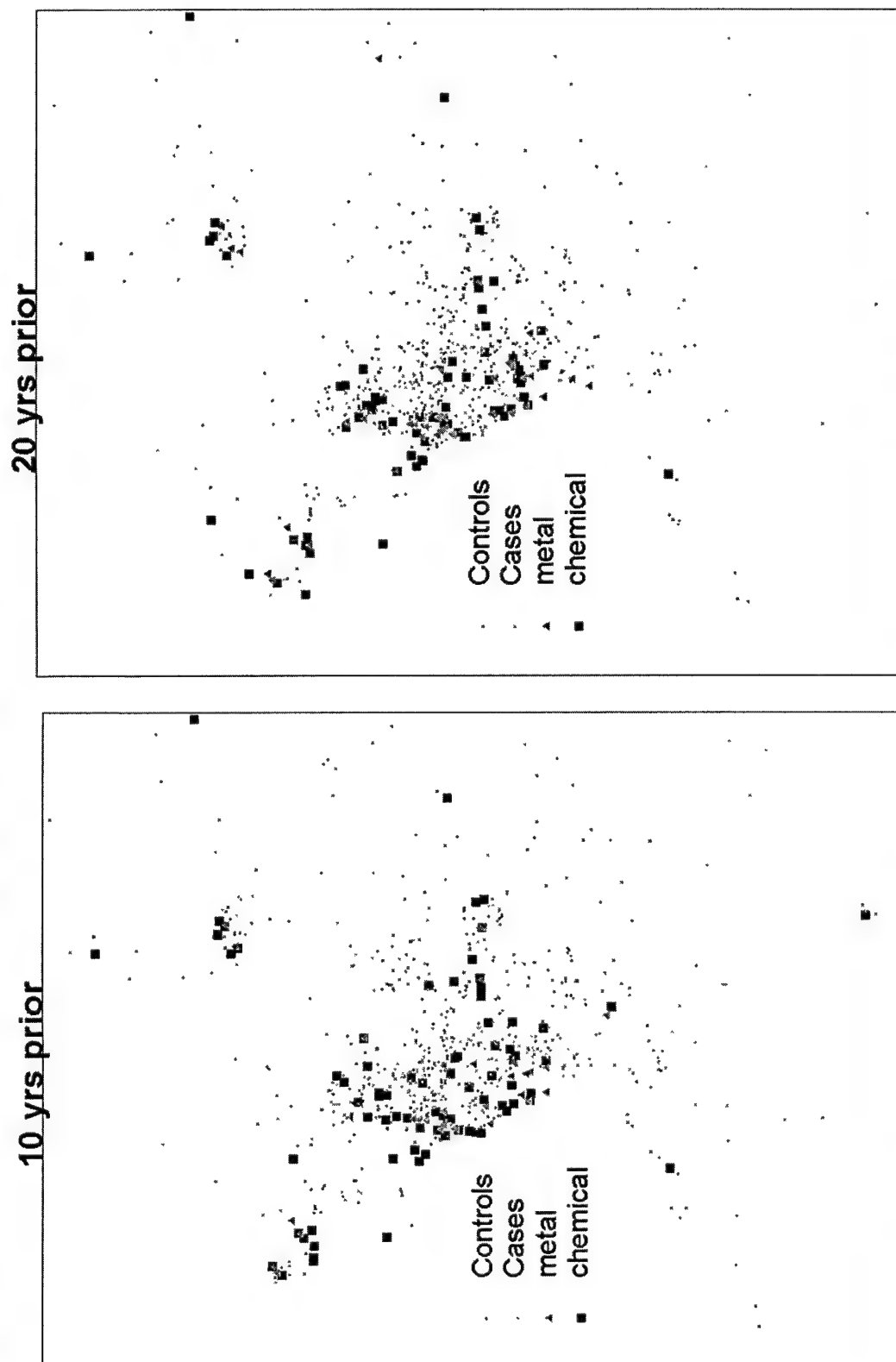
Figure 2. Locations of Industries and Residence by Case-control Status, Postmenopausal Women
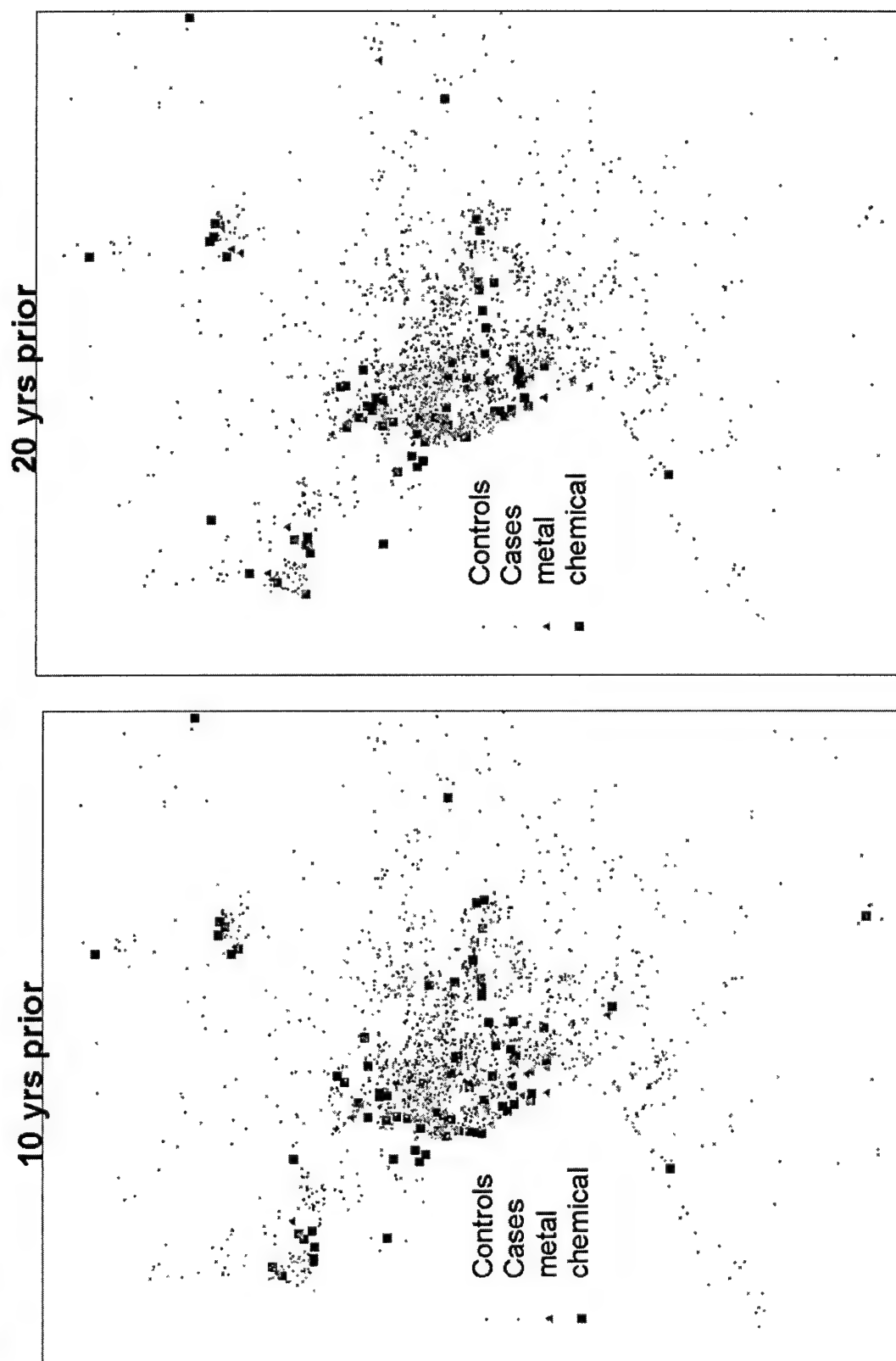

20 yrs prior

.	Controls	
.	Cases	
◄	metal	
■	chemical	

10 yrs prior

.	Controls	
.	Cases	
◄	metal	
■	chemical	

Table 1. Residential proximity to chemical facility 20 yrs ago and Risk of Breast Cancer

Categories of Distance	Cases	Controls	Crude OR	Adjusted OR*	P for trend
				(CI)	
Pre-menopausal					
1: >2.22 miles	61	98	1.00	1.00	
2: 1.16-2.22	35	97	0.58 (0.35-0.96)	0.55 (0.32-0.93)	
3: 0.71-1.16	51	101	0.81 (0.51-1.29)	0.79 (0.48-1.30)	
4: 0.25-0.71	47	86	0.88 (0.54-1.42)	0.83 (0.49-1.40)	
5: <=0.25mile	7	13	0.87 (0.33-2.29)	0.95 (0.34-2.62)	0.06
Post-menopausal					
1: >2.99miles	179	296	1.00	1.00	
2: 1.41-2.99	159	295	0.89 (0.68-1.17)	0.86 (0.65-1.13)	
3: 0.78-1.41	161	291	0.92 (0.70-1.20)	0.91 (0.69-1.20)	
4: 0.25-0.78	143	267	0.89 (0.67-1.17)	0.90 (0.67-1.21)	
5: <=0.25mile	20	35	0.95 (0.53-1.69)	0.84 (0.46-1.54)	0.68

* Odds ratios and 95% confidence intervals adjusted for age, education, race, BMI (1 yr before interview), age at first birth, age at menarche, age at menopause (for post-menopausal women only), number of births, first-degree relative with breast cancer, and previous benign breast disease.

25

Table 2. Residential proximity to chemical facility 10 yrs ago and Risk of Breast Cancer

Categories of Distance	Cases	Controls	Crude OR	Adjusted OR* (CI)	P for trend
Pre-menopausal					
1: >2.53 miles	50	98	1.00	1.00	
2: 1.07-2.53	49	99	0.97 (0.60-1.57)	0.93 (0.56-1.54)	
3: 0.65-1.07	58	98	1.16 (0.73-1.86)	1.25 (0.76-2.06)	
4: 0.25-0.65	35	88	0.78 (0.46-1.31)	0.82 (0.47-1.43)	
5: <=0.25mile	9	12	1.47 (0.58-3.72)	1.72 (0.64-4.59)	0.45
Post-menopausal					
1: >2.50miles	167	296	1.00	1.00	
2: 1.10-2.50	171	295	1.03 (0.79-1.34)	1.05 (0.80-1.39)	
3: 0.65-1.07	157	295	0.94 (0.72-1.24)	0.95 (0.71-1.26)	
4: 0.25-0.65	138	253	0.97 (0.73-1.28)	1.00 (0.75-1.35)	
5: <=0.25mile	29	45	1.14 (0.69-1.89)	1.28 (0.76-2.16)	0.64

* Odds ratios and 95% confidence intervals adjusted for age, education, race, BMI (1 yr before interview), age at first birth, age at menarche, age at menopause (for post-menopausal women only), number of births, first-degree relative with breast cancer, and previous benign breast disease.

Table 3. Residential proximity to primary metal industry 20 yrs ago and Risk of Breast Cancer

Categories of Distance	Cases	Controls	Crude OR	Adjusted OR* (CI)	P for trend
Pre-menopausal					
1: >2.95 miles	61	98	1.00	1.00	
2: 1.62-2.95	48	99	0.78 (0.49-1.25)	0.77 (0.47-1.26)	
3: 0.91-1.62	51	99	0.83 (0.52-1.32)	0.82 (0.50-1.34)	
4: 0.25-0.91	39	87	0.72 (0.44-1.18)	0.74 (0.43-1.26)	
5: <=0.25mile	2	12	0.27 (0.06-1.24)	0.22 (0.05-1.06)	0.14
Post-menopausal					
1: >3.61miles	167	296	1.00	1.00	
2: 2.02-3.61	178	294	1.07 (0.82-1.40)	0.99 (0.75-1.31)	
3: 1.16-2.02	154	292	0.94 (0.71-1.23)	0.96 (0.72-1.28)	
4: 0.25-1.16	151	286	0.94 (0.71-1.23)	0.90 (0.67-1.20)	
5: <=0.25mile	12	16	1.33 (0.61-2.88)	1.45 (0.65-3.23)	0.62

* Odds ratios and 95% confidence intervals adjusted for age, education, race, BMI (1 yr before interview), age at first birth, age at menarche, age at menopause (for post-menopausal women only), number of births, first-degree relative with breast cancer, and previous benign breast disease.

Table 4. Residential proximity to primary metal industry 10 yrs ago and Risk of Breast Cancer

Categories of Distance	Cases	Controls	Crude OR	Adjusted OR*	P for trend
				(CI)	
Pre-menopausal					
1: >3.80 miles	46	98	1.00	1.00	
2: 1.66-3.80	59	99	1.27 (0.79-2.04)	1.37 (0.83-2.26)	
3: 1.06-1.66	51	98	1.11 (0.68-1.80)	1.06 (0.63-1.78)	
4: 0.25-1.06	43	93	0.99 (0.60-1.63)	1.07 (0.62-1.85)	
5: <=0.25mile	2	7	0.61 (0.12-3.05)	0.49 (0.09-2.58)	0.97
Post-menopausal					
1: >3.59miles	164	295	1.00	1.00	
2: 2.02-3.59	164	295	1.00 (0.76-1.31)	0.94 (0.71-1.24)	
3: 1.12-2.02	171	297	1.04 (0.79-1.35)	1.02 (0.77-1.35)	
4: 0.25-1.12	152	283	0.97 (0.73-1.27)	0.99 (0.73-1.33)	
5: <=0.25mile	11	14	1.41 (0.63-3.19)	1.69 (0.73-3.90)	0.41

* Odds ratios and 95% confidence intervals adjusted for age, education, race, BMI (1 yr before interview), age at first birth, age at menarche, age at menopause (for post-menopausal women only), number of births, first-degree relative with breast cancer, and previous benign breast disease.

Task 2: To examine estimated exposure to benzene and to PAHs as a risk factor for pre- and postmenopausal breast cancer, with control for appropriate confounders.

Task 2 is complete. We have geocoded industrial sites, especially major PAH emitters for several decades and will continue with this work for the other relevant periods. We are also working on the development of geographic models to estimate historical PAH exposure from traffic and industrial sites.

We have completed data analysis examining early life exposure to total suspended particulates and exposure to environmental tobacco smoke in relation to risk of breast cancer in adult life. We are interested in both of these exposures as proxies for exposure to PAHs and benzene. Two manuscripts for these analyses have been submitted. Additional publications are being prepared.

A. Early Life Exposure to PAHs and Breast Cancer risk.

A paper, "Breast Cancer Risk and Exposure in Early Life to PolycyclicAromatic Hydrocarbons Using Total Suspended Particulates as a Proxy Measure" has been accepted for publication. A copy of the complete manuscript is included in APPENDIX III.

A paper, "Environmental Tobacco Smoke Exposure in Early Life and the Risk of Breast Cancer" has been submitted for publication. A complete copy of the manuscript is included in APPENDIX IV.

B. Validation of the Traffic Model

Traffic emission is a big component of air pollution and a major source of PAHs in urban areas. While many geographic models have been used to estimate the traffic related air pollution, and to examine its relationship to disease risk, only very few of them were specifically designed to address the PAH exposure. Further, the validity of the model and its ease of application in epidemiologic studies is crucial. Recently, Dr. Jan Beyea and colleagues developed a traffic PAH model for Long Island Breast Cancer project. The model was validated using both spatial and temporal data collected on Long Island. The data used for validation and calibration included PAH measurements carried out on a subset of study subjects, such as soil and carpet PAH concentrations, and PAH-DNA adducts in study subjects' blood. Measurements of carbon monoxide (CO) at an EPA monitoring station were also utilized, because a PAH traffic model can also predict relative CO concentrations by a simple choice of model parameters. They concluded that this model was a valid tool to reconstruct historical PAH exposure.

Although the model parameters determined from Long Island ideally should be suitable for use in other areas, caution still needs to be taken before we can apply it to a different geographic location and study settings. While Dr. Beyea's traffic model was originally designed to calculate the cumulative PAH exposure for Long Island study, it has been modified to estimate the average PAH exposure in a slice of time, i.e. each critical time periods, for our study purpose. To examine if this model is valid in our study areas and also to further calibrate the model parameters, we performed this additional validation study using data from Erie and Niagara area.

Methods

Two sources of data were utilized to conduct the validation study, i.e. measured historical benzopyrene (BaP) air data, and existing carbon monoxide data. The measured BaP data of 12 locations were taken by the New York State Department of Environmental Conservation (NYSDEC) from November 1973 to November 1974, using thin layer chromatography followed by fluorimetric analysis. Hourly CO data was collected through the USEPA/NYSDEC for two monitoring stations in Erie County and two in Niagara county.

All four CO sites were visited and photographed to help validate the latitudes and longitudes of the CO monitors listed on EPA and DEC websites. Information collected during each site visit, e.g. distance from the monitoring station to the road, height of the probe, building and tree structures in the vicinity of the stations, also helped us to better understand the associations between the model predicted and measured CO level. Photographs of the monitor at Site 0016, which had been removed prior to our investigations, were obtained from NYSDEC. Location of the trailer on Site 2006, which had also been moved, was obtained from a local residence. Since there is no traffic count data for the streets close to this monitor, its exact location is not critical. The trailers for the remaining two monitoring sites are still in place. The test year was taken to be 1993, which was the earliest year for which we had reasonably complete data for most of the monitors. For the station at Site 2008, the earliest year when data was collected was 1999.

Traffic count data, was obtained from GBNRTC and NYDOT, and interpolation was used to estimate the data of missing years as described in the traffic model section.

We also collected 24 hour traffic count data from NYDOT for 5 sites (AVC sites) where traffic data was continuously collected throughout the day using inductive loop and axle sensors. Based on this information, the 24 hours traffic curves, i.e. AVC curves, which show the variation in traffic flow over the day were calculated. Hourly emissions of PAH and CO per km of road are taken as proportional to the AVC curves.

To fit the hourly traffic data for each CO site, we used the nearest AVC curve to estimate traffic flow. In cases, where two AVC stations were at comparable distances, we averaged the two curves. Since the model predicted CO is only a relative value, both measurements and predictions were normalized to the values for the unobstructed, Site 0005, by taking the ratio. Since we did not have overlapping years of data for Site 2008, we took 1993's measurement to be the same as the 1999's of 3.8 ppm, based on the fact that changes in the values at the nearby site, 2006, did not show a statistically significant trend from 1993-1996 (p=0.34).

Meteorological data for study areas was obtained from the National Climatic Data Center (NCDC 1999; NCDC 2003).

The traffic model makes use of additional parameters, such as PAH deposition velocity, photodecay rates, and washout rates that were optimized using Long Island data.

We compared the measured BaP data with the model predictions, using correlation statistics and graphs. For CO data, 2 comparisons were made: 1) to compare the annual average CO data against model predictions; 2) to compare hourly predictions of the CO concentrations to the measured values, averaged over an entire year.

Results

Most of the BaP sites were in a city and near a major industry, while only one site, i.e. Site 11, was far from city (Figure 1). The BaP measurements were taken throughout Nov. 1973 to Nov. 1974, with variety of number of sample collected, ranging from 2 to 79 (Table 1).

Figure 2 shows the locations of the CO monitoring stations and the AVC hourly traffic count measurement stations. Also shown are the roads for which traffic count information exists.

Figure 3 shows the range of variation of the hourly traffic count from site to site. Generally, the curves show morning and early evening peaks corresponding to rush hour traffic. One curve, i.e. Site 5383, shows a midday peak, as well.

The characteristics of the 4 CO monitoring stations as well as the data collected were summarized in Table 2. Site 0016 is surrounded by large buildings; all the other CO sites are unobstructed.

Table 3 shows the correlation between model predicted BaP and measured air BaP, and between model predicted and measured hourly CO level. The correlation between measured and predicted BaP was 0.54 (P=0.07); the correlations between the measured and predicted hourly CO ranged from 0.31 to 0.79 for different sites.

Table 4 shows the annual average CO measurement and prediction with or without a constant background CO term added in the model. These data have also been shown on graphs (Figure 4 and 5). Figure 4 indicates that the model under-predicts, particularly at Site 2006, which had missing count data in the immediate vicinity of the monitor. The nearest roads with traffic counts for this site are about 250 meters from the monitor. To move this point up to the measurement, the average CO concentration would

31

have to be doubled, which provides an estimate of the error rate that might be occurring at residences far from major roads. For the Site 2008, although there were missing data for one of the nearest streets, we did have traffic data for a side street that is quite close, resulting in little under-estimation of the measured value.

Discussion

Despite large variations about the regression line, the predicted BaP air concentrations correlated with the measured concentrations and the relationship between the two is approximately linear, as expected (Figure 6). The correlation is modest (0.54), which is not too surprising, given the limited information about the air measurements concerning exact sample locations, time of day and date of the measurements, varying number of samples per location. Also, some of the PAH detectors were deliberately placed in the vicinity of industrial emitters, which may have increased the variance of the data compared to more typical locations in the area. Nevertheless, the large variations from the regression line raise the possibility of large uncertainty in predicting PAH air exposures that could serve to mute the ability to find a dose response, if one exists.

Although the traffic model was designed to predict traffic-related PAH, it can also predict the relative concentrations of CO from traffic. The CO model is equivalent to running the PAH model with certain parameters set to zero, namely "deposition velocity," "light decay," and "rain washout." By comparing the relative CO exposure, predicted vs. measured, we thus check the dispersion part of the model, but not the removal processes.

Using graphs, Figure 7 to 10, we compared hourly predictions of the CO concentration to the measured values, averaged over an entire year, for different sites. We show the results with the constant background term added, because it improves the visual appearance of the fit, although it does not change any of the correlation coefficient. Except for the Site 0016, the meteorological dispersion patterns have changed the temporal emission pattern shown in Figure 3, muting the afternoon traffic peak and extending the curves into the late evening and early morning hours.

For the Site 2006, we have no traffic data for the road immediately next to the monitor station. As a result, the distant contributions, with muted morning peak are more important. Would traffic data for the closest streets be added, the morning peak would increase. This gives us an idea of what will happen with residences that are relatively far from roads with traffic counts (Figure 9).

The presence of the strong afternoon peak in the measured data for the Site 0016 and the absence of the extended evening peak, are something of a puzzle (Figure 8). Possibly, traffic at this government center dropped off more rapidly after 6 pm than at the sites at which hourly traffic counts are available. The complex dispersion patterns set up by the tall buildings at this site might have led to enhancement of the afternoon emission peak. In any case, these fine points in the shape of the temporal pattern have little to do with the annual average values that are of interest for the epidemiological aspects of the study. In particular, correlation coefficients for hourly patterns do not reflect the average values, which are of interest for the epidemiologic study. Still, it is interesting to see that, in general, the meteorological dispersion model, which is adding up contributions from both local and distant streets and disbursing them differently at different times of the day, produces output that approaches the measured time trend, albeit with varying success.

When we added a constant background term into the traffic model, as shown in Figure 5, brought the predictions in better agreement with the measurements, although there was still some under prediction at site 2006 (note that it still under-estimated the CO by about 25%). It is possible that contributions from roads throughout the area where traffic counts are not available, along with industrial emissions, may be contributing an overall background level of CO. Residential heating is unlikely to be making a significant contribution, because examination of the CO data by season shows no substantial winter increase that might have occurred from space heating (Figure 11).

In summary, nothing in our validation study using local data suggested that there need be any significant modifications to the traffic model, as calibrated with the more extensive Long Island data. Based on this study, it is appropriate to use the model, with site specific meteorology and traffic count data, to reconstruct the historical PAH levels in Erie and Niagara counties, although cautions need to be taken for people lived far away from major roads with traffic count.

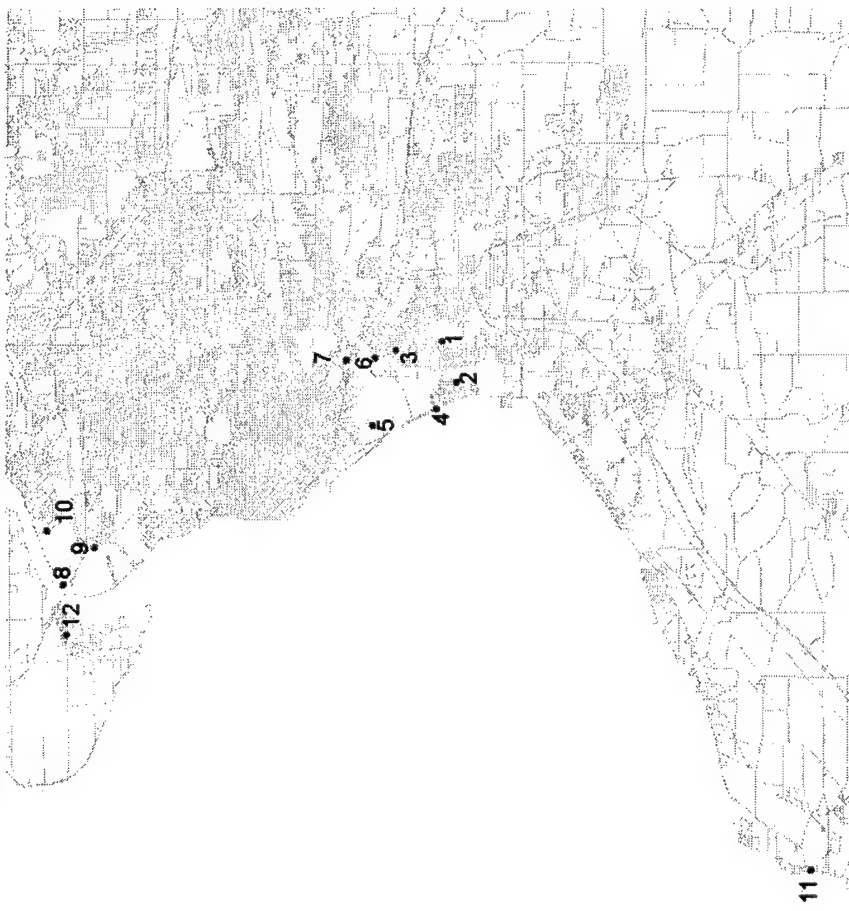Figure 1. Locations of the 12 BaP Sites

Table 1. BaP Data for 12 locations in Erie and Niagara County, Nov. 1973 - Nov. 1974

Site id	Place of sample colletection	Address	City	# of samples	BaP (ug/M^3) Average	Max.	Min.
1	OLVH (our lady of victory hospital)	800 Ridge Rd.	Lackawanna	2	0.01621	0.0305	0.00193
2	Lack. STP (lackawanna sewage treatment plant)	252-282 Lehigh St.	Lackawanna	35	0.01251	0.0487	0.00057
3	HF School (holy family school)	South Park & Tifft St. (920 Tifft St.)	Buffalo	49	0.012136	0.06843	0.00006
4	PS #4 (Wilson School)	6th St. (36 11th St.)	Lackawanna	10	0.055334	0.19	0.002764
5	Great Lakes Dredge & Dock	Foot of Katherine St.	Buffalo	76	0.00287	0.02384	0.00005
6	PS #28	1515 S. Park Ave.	Buffalo	68	0.009221	0.06556	0.0002
7	PS #26	84 Harrison St.	Buffalo	44	0.007382	0.046219	0.00007
8	Ascension Chemical	River Rd. (400 River Rd.)	Tonawanda	75	0.001275	0.0212	0.00005
9	Thruway Garage	Grand Island Blvd.	Tonawanda	79	0.003751	0.08141	0.00013

10	Ton. CAM (tonawanda sewage treatment plant)	Two Miles Creek Rd. (779 Two Miles Creek Rd.)	Tonawanda	47	0.00222	0.012	0.00001
11	Angola STP (angola sewage treatment plant)	Old Lake Shore Rd.	Angola	47	0.000808	0.02446	0.00004
12	Grand Island (kaegebein elementary school)	Love Rd. & Beaver Island Pkwy. (1690 Love Rd.)	Grand Island	31	0.000442	0.00316	0.00006

Figure 2. Locations of the 5 AVC Sites and 4 CO Monitoring Stations*
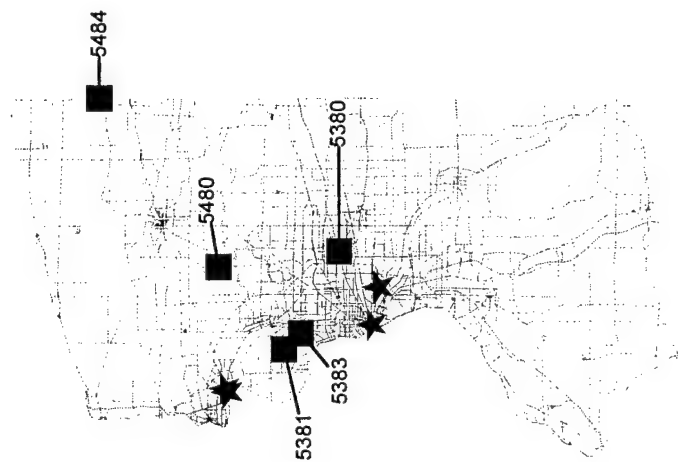
* AVC locations (squares) and CO monitoring stations (stars); the two CO sites in Niagara County are too close together to resolve at the map scale.

Figure 3. Hourly Traffic Count Curves for the 5 AVC Sites
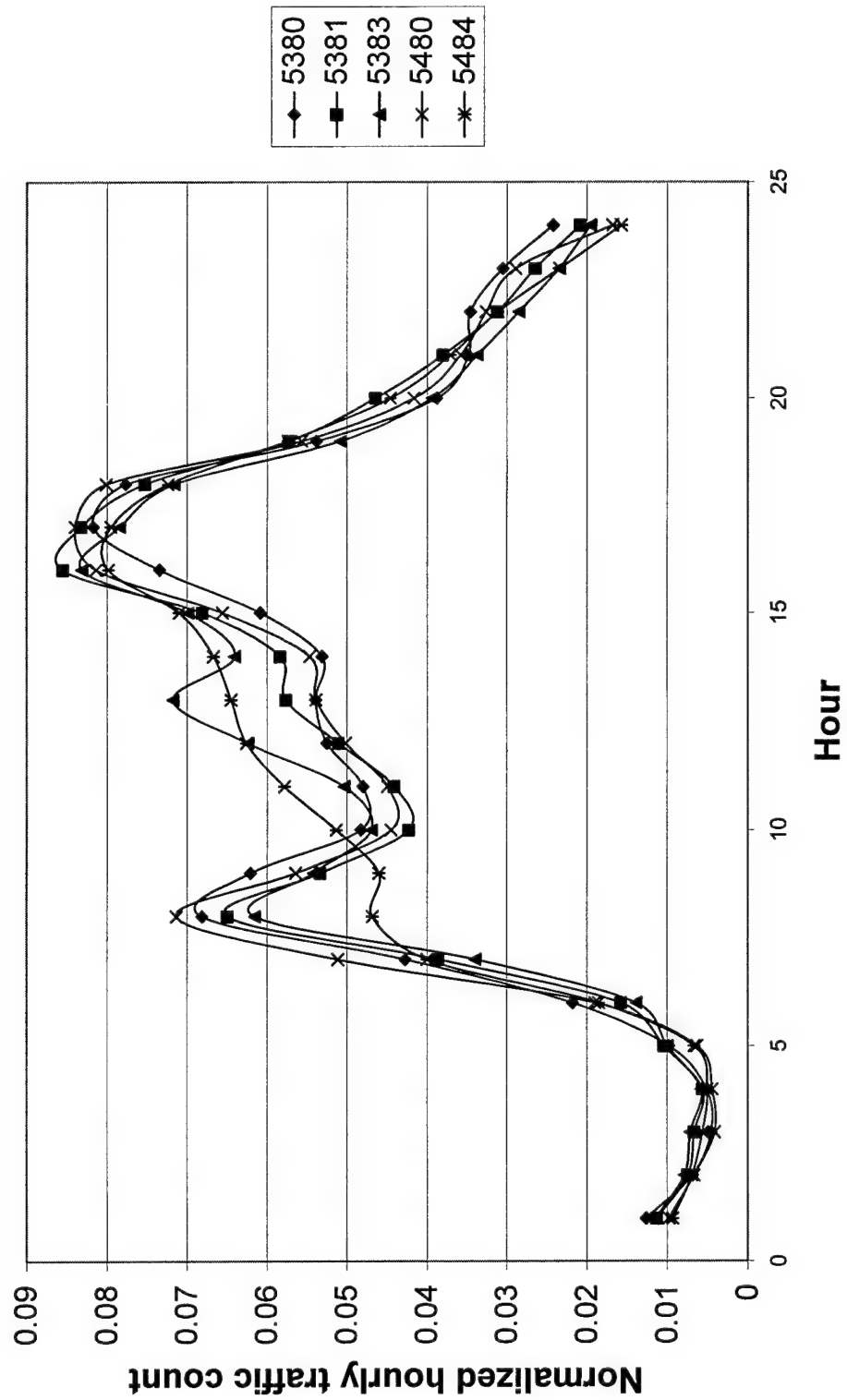
Table 2. Descriptions of the CO Monitoring Stations

Site ID	Hours of valid data (%)	AVC curve used for hourly emissions[a]	Lat/Lon	Still exist?	Site description[b]	Traffic data for closest street(s)?	Location identifiers	Data year
0016	97.8	5380 + 5383	42.886354/ -78.877028	N	Tall buildings in government center. Probe extending from wall.	Y	NYSDEC photos of probe	1993
0005	99.0	5380	42.87688/ -78.8096	Y	Trailer 100 meters from major roads	Y	NYSDEC coordinates & photos	1993
2006	96.5	5381 + 5480	43.08582/ -78.99392	N	Trailer 250 meters from major roads	N	Photos & local resident's memory of trailer location	1993
2008	98.2	5381 + 5480	43.08227/ -79.000644	Y	Trailer	Y	Visit and photos	1999

[a.] Number indicates counter designation. [b.] Probe heights all 4.5 m.

Table 3. Correlation Between Model Predicted and Measured BaP/CO Level

	Pearson Correlation	Spearman Correlation
	(P value)	(P value)
BaP	0.54 (0.07)	0.43 (0.17)
Hourly CO		
Site 0005	0.79 (0.00)	0.72 (0.00)
Site 0016	0.58 (0.00)	0.45 (0.03)
Site 2006	0.46 (0.03)	0.55 (0.01)
Site 2008	0.31 (0.15)	0.23 (0.29)

Table 4. Measured and Predicted CO at 4 different locations in Erie and Niagara County, 1993

Site Id	Annual Average CO (ppm)	Normalized to Site 0005	Predicted CO Ratio (no background)	Predicted CO Ratio (with background)
0005	6.6	1	1	1
0016	8.7	1.32	1.23	1.15
2006	5.0	0.76	0.36	0.57
2008*	3.8	0.56	0.45	0.63

* Using 1999's measurment.

Figure 4. Predicted vs. Measured CO Ratio in Different Sites, with no Backgroup CO Term Added
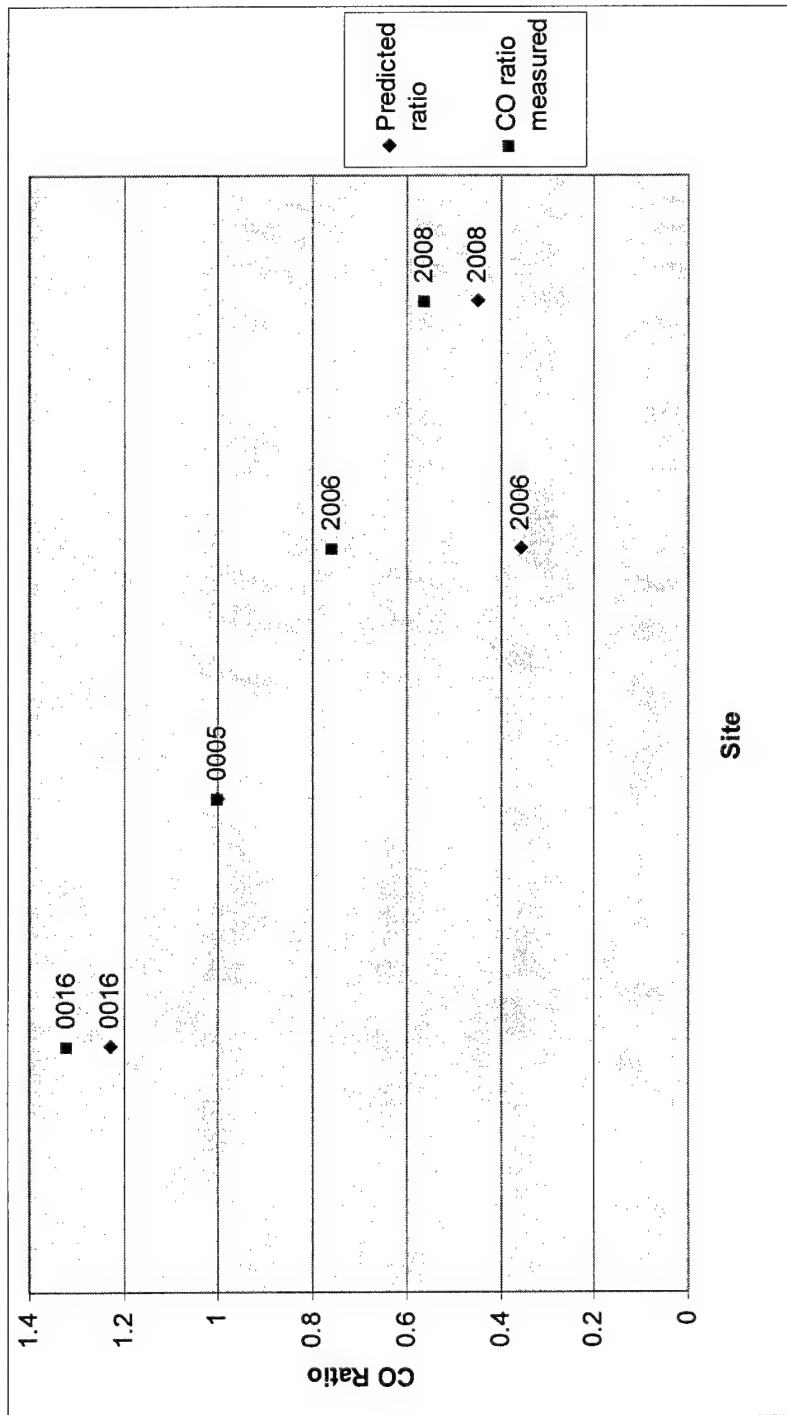
Figure 5. Predicted vs. Measured CO Ratio in Different Sites, with Constant Backgroup CO Term Added
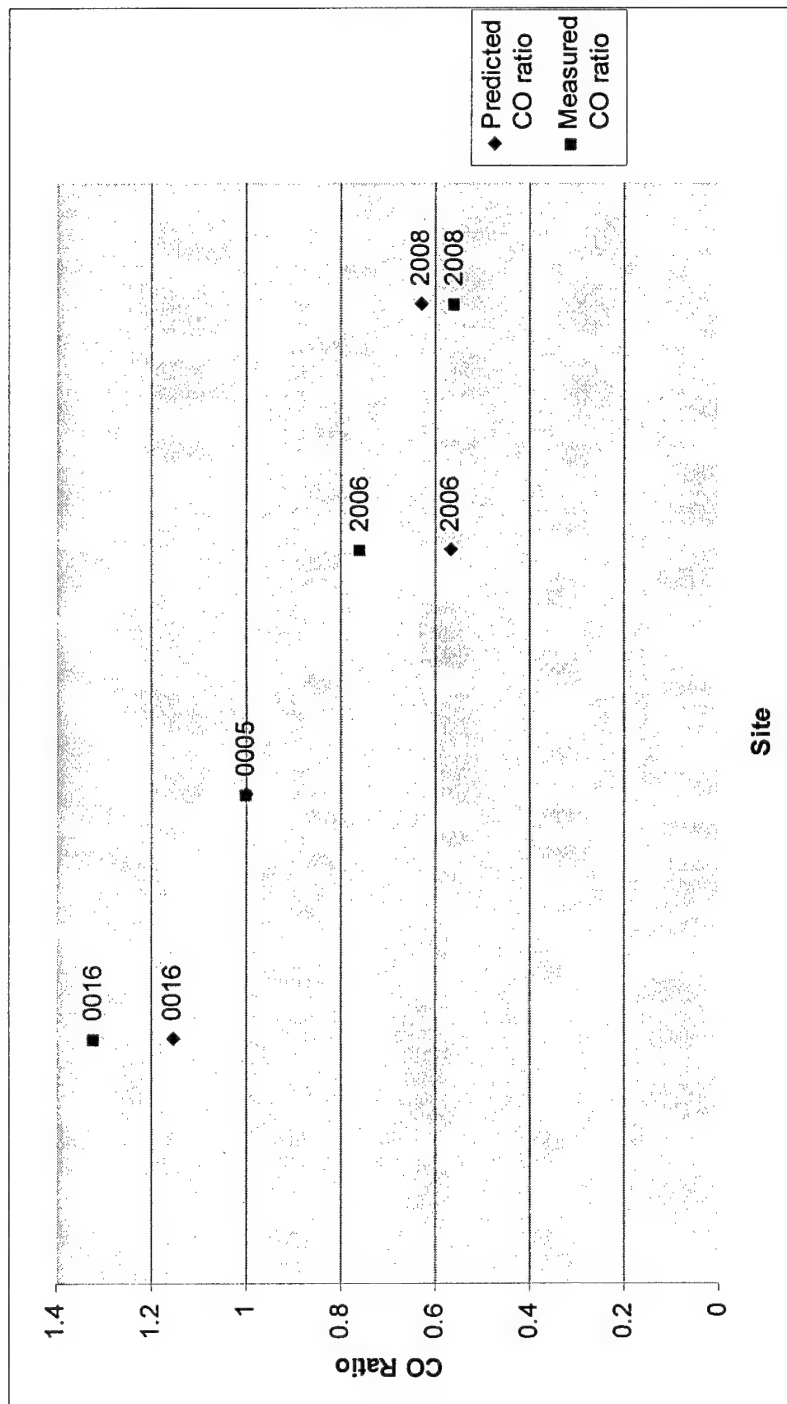
Figure 6. Predicted vs. Measured BaP in 12 locations in Erie and Niagara County (Log transformed)

$y = 0.8884x - 10.579$

Air BaP measurement
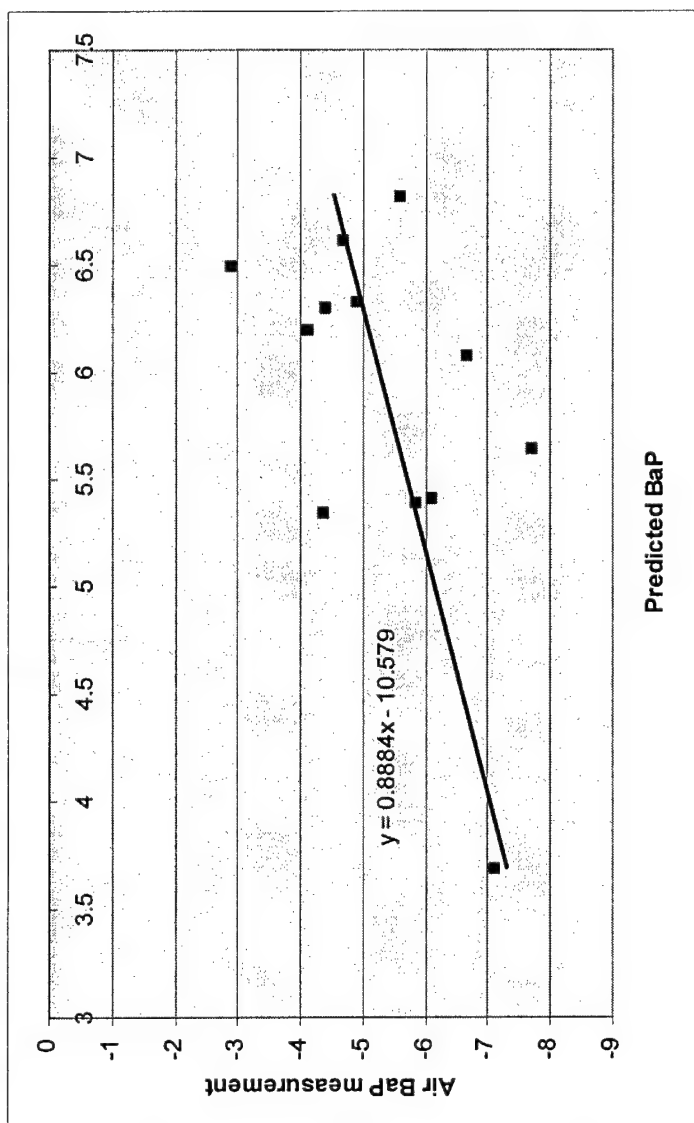
Predicted BaP

Figure 7. Predicted (with Constant Background Added) and Measured Hourly CO, Site 0005
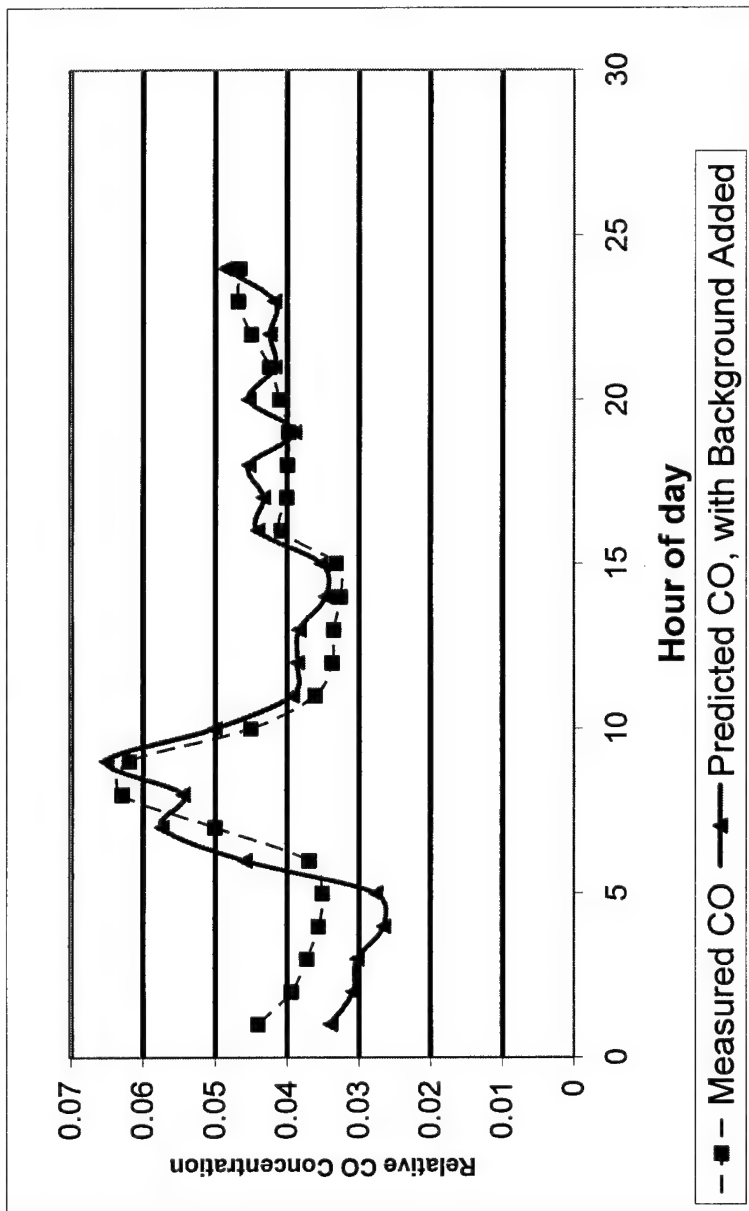
Figure 8. Predicted (with Constant Background Added) and Measured Hourly CO, Site 0016
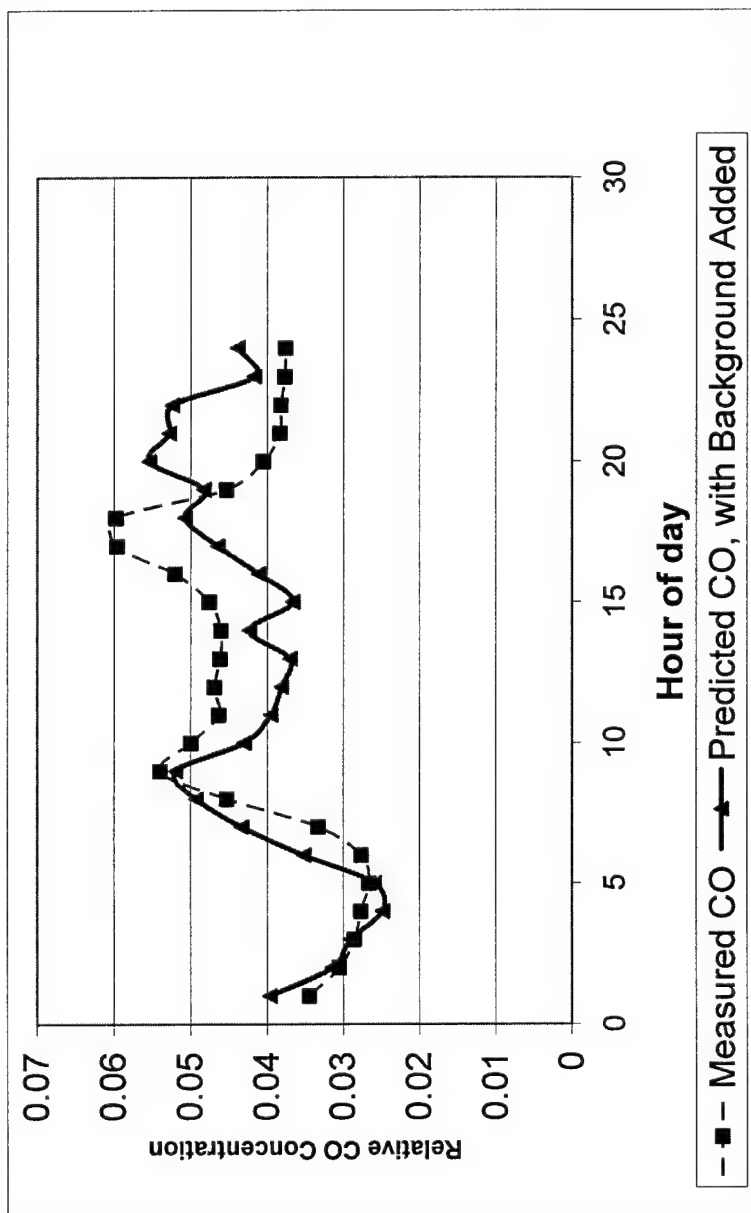
Figure 9. Predicted (with Constant Background Added) and Measured Hourly CO, Site 2006
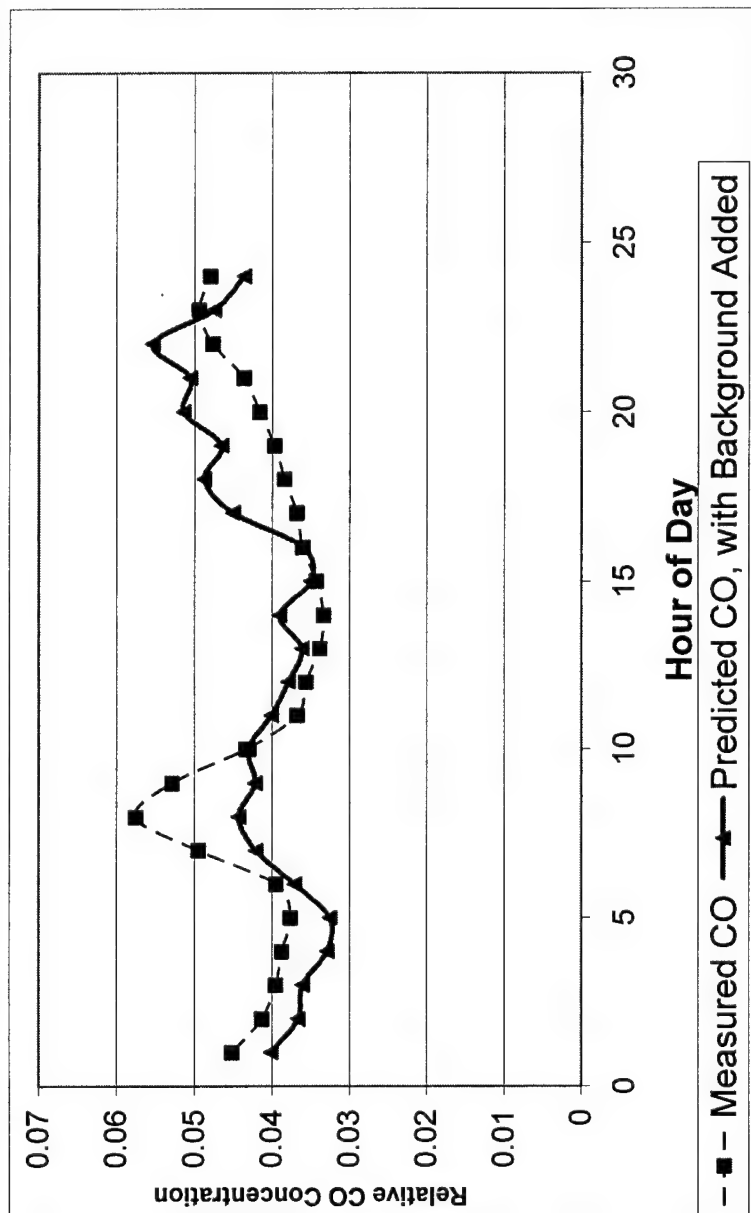
Figure 10. Predicted (with Constant Background Added) and Measured Hourly CO, Site 2008
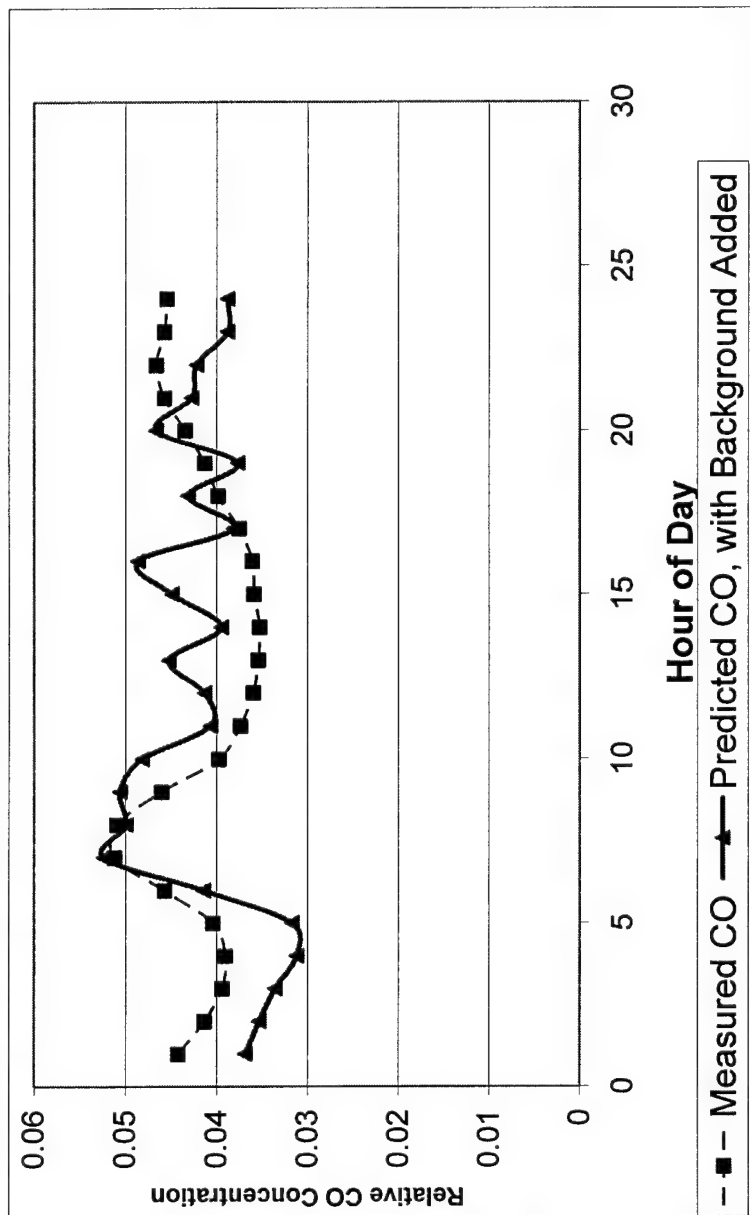
Figure 11. Measured CO at Site 0005 by Season, 1993

C. Traffic PAH Exposure and Risk of Breast Cancer

Traffic is one of the major sources of PAH exposure in cities, especially after the automobile became a common means of transportation. While some studies suggest that food is the predominant source of PAH for human exposure (ATSDR), the estimates of PAHs from diet are less reliable because of the uncertainty of food origins. Study also shows that dietary sources of PAHs are only weakly correlated with internal PAH dose, for example PAH-DNA adduct. According to reports from ATSDR, the major sources of PAH air exposure in the 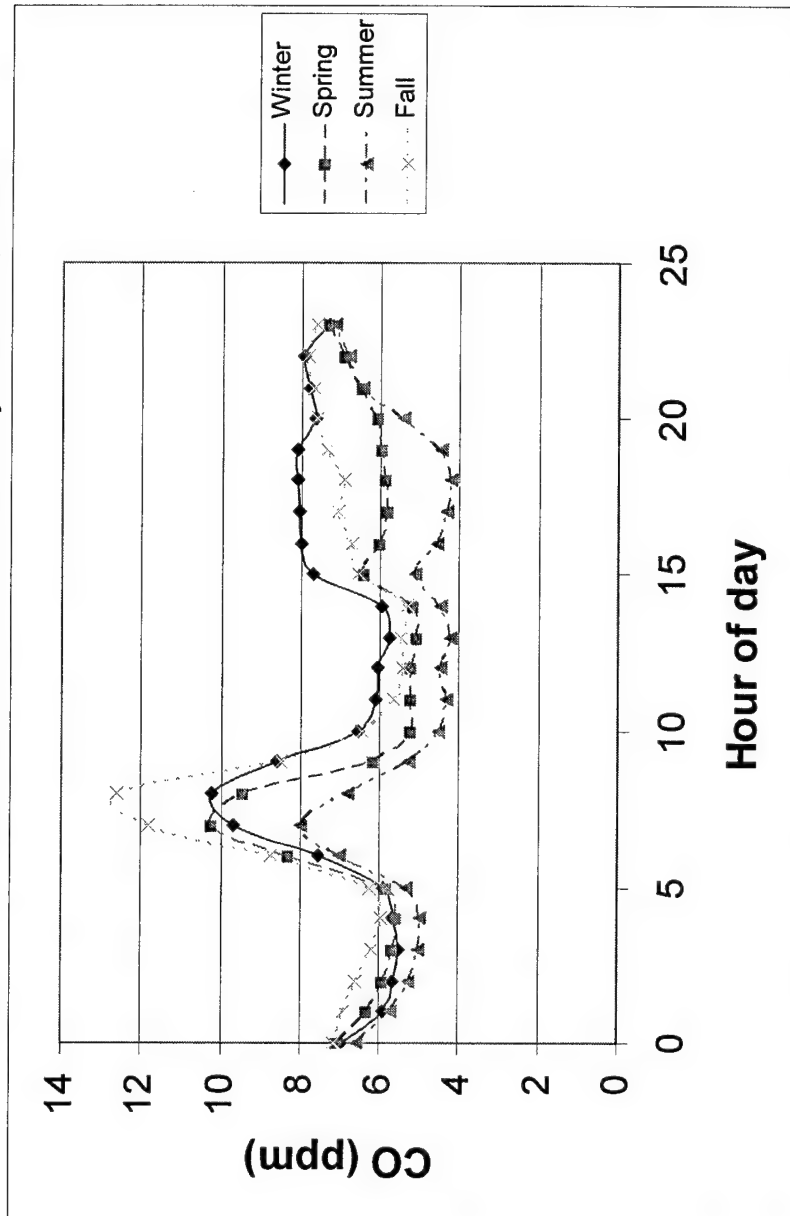U.S. are wood burning and traffic emissions. Mobile emissions account for 20% of the total PAH emissions, and this is similar to reports from 6 European countries. The percentage of traffic emission is even higher in urban areas. One study found that motor vehicles may account for up to 90% of the total particle-bound PAH in the downtown area of a city. This has been confirmed by another study in the UK. Traffic emissions are one of the best characterized important sources of PAHs. The amount of PAHs in traffic emissions is determined by the type and parameters of fuel used, driving conditions, temperature, exhaust treatment, and engine adjustment.

In this study, we used a GIS traffic model, developed by Dr. Jan Beyea and colleagues, to estimate the historical residential exposure to PAHs from traffic. As mentioned in the previous section chapter, this traffic model has been validated as a useful tool to reconstruct historical PAH exposure from residential locations, using data collected from Long-Island Breast Cancer Study, as well as some additional data from our study area. The association between the traffic PAH exposure and the breast cancer risk was examined for time period of potential breast cancer development, i.e. at menarche, first birth, 20 and 10 years prior to cancer diagnosis.

Methods
Data Collection

Data from several sources were collected to determine the PAHs from vehicle emissions.

Historical county traffic volumes were obtained from the Greater Buffalo-Niagara Regional Transportation Council (GBNRTC) for the years from 1971 to 2002, and the New York State Department of Transportation (NYDOT) for the years between 1960 and 1975. In both sources, the traffic volume was recorded for each segment (with a start and end point) which may approximately have similar traffic flow. The length of each segment varied from 0.1 to 10 miles. While the NYDOT data provide us only the data for touring route system, the one from GBNRTC contains also local highway data. Five functional classifications of roads are available in the GBNRTC traffic system, including interstate, expressway, principal arterial, minor arterial, and collector (Figure 1), covering the major roads in the traffic network (Figure 2). Annual Average Daily Traffic (AADT) in both data sources represents the total traffic volume in both directions, taking into consideration of the types of vehicle and seasonality.

Tailpipe emission data were collected from previous journals and reports, including measurements carried out in tunnels or on individual vehicles run in place on test beds. Based on these raw data, two curves, i.e. tunnel fit and scaled dynamometer fit were developed to model the historical tailpipe emission (Figure 3).

Meteorological data were obtained from Environmental Protection Agency (EPA) and the National Climatic Data Center. These data include wind speed, direction, "stability class," or equivalently temperatures at two heights on the meteorological tower, and twice-daily "mixing height".

Exposure Assessment

In our traffic Model, the total traffic PAHs emissions composed three terms, i.e. cruise (warm engine) emissions, cold engine emissions and intersection emissions. Two separate weights, generated from Long Island Breast Cancer Study and the validation results using local data, were applied to the model to adjust for the higher emission of the cold engine and intersection. To obtain the indoor PAH exposure, we applied a building penetration factor (0.75) into the total traffic PAHs emissions.

1) Cruise emissions: Computed as the product of tailpipe emission and traffic counts in the road network, i.e. Emissions=tailpipe emissions per vehicle-km * traffic count * road segment length.

Historical traffic volume information in Erie and Niagara Counties were extracted from the reports of GBNRTC (1971-2002) and NYDOT (1960-1975, State roads only), with AADT as the basic unit of measurement. We then assigned these traffic volumes to each of the 54,494 major road segments in the two study counties, using the nearest available measurement on that road within 10 km, and we repeated this for all the years in the study windows.

Traffic volume was not monitored every year which resulted in gaps in the traffic data. Interpolation or extrapolation (within 10 years) was used to estimate the traffic volume in missing years, when multiple-year traffic data were collected for a point; countywide traffic growth rates were used to fill these gaps, when only one-year traffic data was collected.

In this study, we ignored the effect of PAH exposure from areas surrounding the two study counties. These should add little because the study region is bounded by Lake Erie and suburban area with very low traffic flow.

Since traffic data started in 1960, we did not use logarithmic extrapolation for cruise emissions before that time, due to a concern regarding misclassification. Similarly, this same rule was applied for the calculation of cold engine and intersection emissions.

2) Cold engine emissions: Cold engine emissions usually contain higher levels of PAHs than warm engine emissions, thus we calculated them separately. Similar to warm engine, we collected historical AADT and tailpipe emissions to construct the cold engine emissions.

AADT was first calculated in the census block level, estimated as the product of total number of cold starts per household per day and the number of households in each census block, and then was assigned to the roads within the census block. We obtained the number of cold starts in 1995 from the Nationwide Personal Transportation Survey (NPTS 1995), and estimated the number of cold starts between 1960 to 1995 by scaling from the national figures from Hu. The number of households was collected from the historical US census data. We assumed that cold engine emissions would last for 1km travel distance. Once the AADT in each census block was calculated, we then assigned them among the roads, using the inverse square distance to them from the centroid of the census block. We set 1km as the total trip length from the center of the centroid to the last

51

point included on a major road, and we attempted to assign the emissions uniformly to all local residential street segments lying with a census block.

3) Intersection emissions: Since accelerating and decelerating in the intersection may increase emissions, we assigned a weight to the segments within 200 meters of an intersection. Since there was no detailed information about the traffic control at intersections, we assumed that 10% of the traffic was exiting or entering, thus emitting more PAHs.

Statistical Analysis

All analyses were done separately for pre- and post-menopausal women unless otherwise specified. In our study, women were defined as post-menopausal using specific criteria obtained during the personal interview. We will take into account: hormone use (present or past), reason for cessation of menses (natural, surgical, chemotherapy, radiation, etc) and age. All women 60 years of age were considered post-menopausal.

To describe the distribution of the studied variables, means and standard deviations (SDs) were presented for the continuous variables between cases and controls groups, and T-tests were used to compare means. Chi-square tests were used for categorical variables. Pearson correlations were presented among the continuous variables to examine the interdependency of these variables.

Unconditional Logistic regression were used to calculate the odds ratios (OR) and 95% confidence intervals (CI). As the linear dose response relationship is assumed between PAH exposure and breast cancer risk, we entered the PAH estimates from the geographic models as a continuous variable into the regression model, to test the linear trend. The common breast cancer risk factors as potential confounding variables were adjusted in the model, including age, education, race, BMI, age at first birth, age at menarche, age at menopause (for post-menopausal women only), number of births, first-degree relative with breast cancer, previous benign breast disease. To test the potential effect modification, the product of PAHs estimates and smoking/diet were introduced into the regression model, and stratified analyses were performed. Stratified analyses by ER/PR status were also conducted.

Due to the fact that PAH exposure estimates calculated in this study showed a skewed distribution, natural log was taken for all PAH exposure estimates to make it normally distributed. All the PAH exposure estimates mentioned below therefore are the ones taken natural log, unless otherwise specified.

Results

The characteristics of breast cancer cases and controls by menopausal status are listed in Table 1-1 and Table 1-2. Although in our study, controls were matched to cases by age and race, small differences still existed which might due to the fact of only frequency matching and/or only small percentage of the study participants included in each time period analyses. For example, among post-menopausal women in menarche analyses, cases were on average 2 years older than controls (49.4 vs. 47.3 years old), noticing only 128 women (i.e. 52 cases and 76 controls) were included in these comparisons; among post-menopausal women in 20 years prior analyses, there were more whites in cases than in controls, 93.3% vs. 90.4%. The distribution of education and BMI were similar between cases and controls.

Correlations between traffic model predicted PAHs and the control variables are listed in Table 2-1 to Table 2-4. Using different parameter settings, which includes Long Island/Buffalo intersection weights and with/without problematic traffic counts included, we obtained four sets of PAH estimates, i.e. BaP (R5) to BaP (R8). For all time periods and for both pre- and post-menopausal women, the corrections among them are very high. The nonparametric correlations among these different estimates are even stronger, indicating that most of the subjects will fall into the same quartile category in all following regression analyses no matter which estimates were selected. Thus we decided to use BaP (R6), where the Buffalo intersection weights were used and problematic traffic counts information were excluded, as the only PAH estimate in all subsequent analyses. Distance to the closest road with traffic count and year at exposure were negatively correlated with PAH estimates, while traffic count in 1990 was positively correlated with PAH estimates. This showed that the PAH estimates were a combination effects of distance to closest road, year at exposure and traffic count. Worth mention that year at interview was negatively correlated with PAH estimates. Due to the fact that the cases were generally interviewed later than controls, year at interview was considered as a controlled variable in all subsequent regression analyses. Further, all other correlations were not strong.

The correlations within the control variables were not strong. Age was generally positively correlated with age at menarche, age at menopause and parity; and BMI negatively correlated with years of education and age at menarche.

The crude and adjusted ORs are shown in Table 3. For the crude comparison, compared to controls, the cases were more likely to have a higher PAH exposure at menarche and at women' first birth. At menarche, among pre-menopausal women, OR is 1.99 for the highest quartile. At first birth, among post-menopausal women, OR is 1.61 in the highest quartile. However, this increased risk was not seen among the post-menopausal women in the menarche analyses and the pre-menopausal women in first birth analyses. In consideration of the fact that controls were generally interviewed earlier than controls, and air PAH dropped through the years, year of interview were entered into the adjusted models. After controlling for year at interview as well as other factors, i.e. age, education, race, BMI, age at first birth, age at menarche, age at menopause, number of births, first-degree relative with breast cancer, previous benign breast disease, the suggested increased risk at menarche and first birth analyses still existed. Traffic PAH estimates were also entered into the regression models as a continuous variable to examine if there was a potential dose-response effect between PAH exposure and breast cancer risk. We found, only among pre-menopausal women at menarche analyses, the P for trend was significant ($p=0.03$).

To deal with the problem that the traffic model may underestimate the PAH exposure for women lived far from a road with traffic counts, suggested by the traffic model validation results mentioned in previous chapter, we restricted the analyses to women living within 250 meters from a road with traffic counts, the primary findings remain unchanged (Table 4).

To better explore the potential confounding effect of social economic status in our study, we tried adjusting income instead of education (or both income and education) in the analyses. The results showed that using these different sets of variables in adjustment did not significantly change our results (Table 5 and 6).

When we stratified the subjects by smoking status, the increased risk was only seen in non-smokers (Table 9 and 10). Among pre-menopausal women at menarche analyses, cases were about 7 times more likely to be in the highest quartile group than controls. We found increased risk of breast cancer for both pre- and post-menopausal women at the first birth analyses (Table 9).

To examine if there is any confounding effect of smoking, we entered packyear of smoking in the regression model. Controlling for smoking did not change the previous findings (Table 11).

Effect modification by ER/PR status was examined. In these analyses, the whole set of controls were included in all regressions. As indicated in Table 12 and 13, the results were similar for both ER(-) and ER(+) strata, and there was no increased breast cancer risk in the higher PAH exposure categories. Stratification analyses by PR status showed the similar results, except for at first birth analyses, where the PR(+) analyses suggested a little stronger increased breast cancer risk among women exposed to higher PAHs, comparing to PR(-) (Table 14 and 15).

Discussion

Studies have found exposed to traffic emissions associated with increased risk of total childhood cancer and childhood leukemia, but to our knowledge, this is one of the first studies that examined traffic emission and breast cancer risk.

Using a GIS traffic model to estimate the residential exposure to PAHs from traffic in potential critical time periods of breast cancer development, i.e. at menarche and at first birth, our study also indicated that women exposed to higher level of traffic PAHs were associated with higher risk of breast cancer, and this increased risk was only existed in non-smokers.

The fact that increased breast cancer risk was only observed in earlier life time period, i.e. at menarche and first birth, may suggest that the observed association is due to the general high level of PAH exposure in earlier years. Additional analyses by years of exposure were conducted to clarify this. Five years were selected, namely 1960, 1965, 1970, 1980, and 1990. Only addresses that the participants resided in during each of these 5 years were included in each subset analysis. The increased risk was found only in the years 1960, 1965 and 1970 analyses among pre-menopausal women, but not post-menopausal women (Table 16). These findings may also provide additional evidence to our hypothesis of the importance of earlier life exposure.

The striking difference for analyses stratified by smoking status is quite interesting. Although smoking is another important source of PAHs, most previous studies did not find an association between smoking and breast cancer risk. Our study suggest that among smokers whose PAH exposure may be already high, additional exposure from traffic may not make a different in term of breast cancer risk. However, among non-smoker whose PAH exposure from smoking are potentially low, increased traffic exposure may greatly increase breast cancer risk.

Study has suggested that in situ cancer may be more sensitive to environmental exposure with geographic variation. Although there seem to be some indications of this, our results are limited by the small sample size (Table 7-1 and 8). Since we only have very few women with in-situ breast cancer in our study, we do not really have enough

power to draw any conclusions regarding this, even after we tried to adjust fewer covariates in reduced models (Table 7-2).

Although our model is novel way of using a geographic traffic model to reconstruct the historical PAH exposure, this method has many limitations. 1) Row housing: Since the percentage of row housing is small in the study area, we did not incorporate "canyon effects" into the modeling. 2) Historical changes in road network: In spite of the roads changes over time, we did not include an algorithm to "remove" road sections from the network in backwards extrapolation. This means that in a few cases, we might overestimate exposures in the early years due to the inclusion of emissions from roads not yet built. 3) In-vehicle exposures: We did not account for exposure to traffic PAHs while driving. 4) Neglect of other PAHs air exposures: Our study is premised on the experimental finding that traffic exposures are a dominant source of airborne PAHs, both in and outdoors. Thus, although we neglected emissions from indoor sources (e.g., PAHs in airborne combustion products of cigarette smoke and cooking), we expect that the amount of PAHs from traffic alone is sufficient to allow its detection in breast cancer risk, if there does exist an association between breast cancer and the PAHs delivered on ultra fine traffic particulates. 5) Limited data on earlier exposure: Due to the limited year of the traffic count information that was available, we were not able to include resident addresses of prior to 1960 in this study. For this reason, we were not able to examine the association between PAHs at birth and breast cancer (noting that majority of the participants were born before 1960), although exposure at birth is another very important critical time window in term of development of breast cancer. For the same reason, the sample size for both menarche and first birth analyses was greatly reduced.

As a case-control study design, this study may be prone to different types of bias. Selection bias may be a concern especially as we have relative low participant rate. However, as cases living near our study center were more likely to participate, we found the same pattern in controls. And the other common breast cancer risk factors were not different between subjects included and excluded, except that people exclude is a little older, and more likely to have 3 or more children. Recall bias is not likely happen in this study. At the time of study, participants were generally unaware about our study hypotheses. Misclassification do exists from the self-report residence, as well as from using traffic model to estimate PAH exposure, however, it's more likely to be non-differential and may draw study results to null. As we indicated in the previous chapter of model validation, there was a potential problem of underestimating exposure for people live far away from major road with traffic counts. However, restricting people to only those who lived close did not change our primary findings.

In this study, most of the study participants provided their lifetime residential history; and significant efforts have been made to reconstruct historical PAH exposure. This allowed us to examine the relationship between breast cancer and PAHs in different time windows, especially those critical time periods of breast tissue development.

Our study population was a residentially stable population. While the mean age of our subjects was about 56 years, people were in average moving 5 times lifetime. The stable of this population gave us unique opportunity to examine the long-term effect of PAHs, and made examining in only critical time windows more meaningful. Subjects

were likely to have been staying in the same residence for many years before the time period we examined.

Further, common risk factors for breast cancer and other characteristics of the study participants have been collected by interviews and self-administered questionnaires. This made it possible to examine the potential effect of confounders and effect-modifiers in the models. We also have some data on other sources of PAHs, such as smoking, which allowed us to examine and adjust their effects in the traffic models.

In summary, using a geographic model, we found some evidence of high traffic PAH exposure associated with increased risk of breast cancer, particularly for the earlier time exposure. This study provides additional evidence to the hypothesized association between PAHs and breast cancer. However, in interpretation of the study results, there must be caution because of study limitations mentioned earlier. Future studies are need, for both critical time windows and lifetime exposure, to confirm our findings in this study.

Figure 1. Functional Classification of the Traffic System in GBNRTC

Functional Classification
COLLECTOR
EXPRESSWAY
INTERSTATE
MINOR ARTERIAL
PRINCIPAL ARTERIAL

57
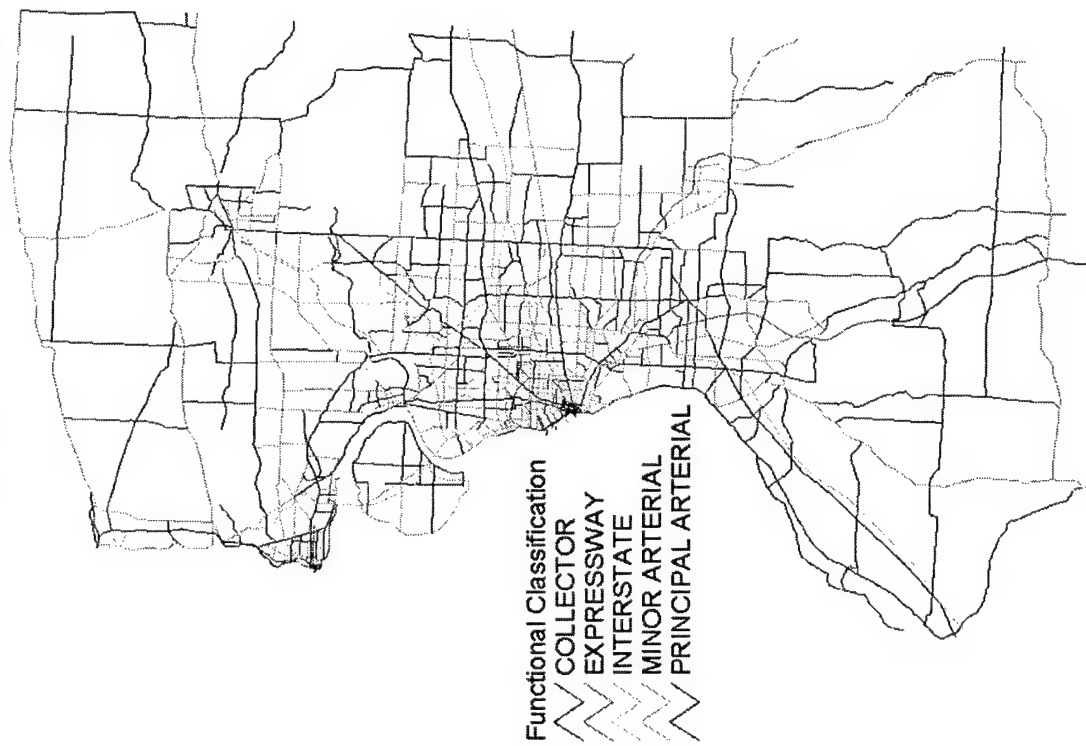
Figure 2. Density of the Roads with Traffic Count Information
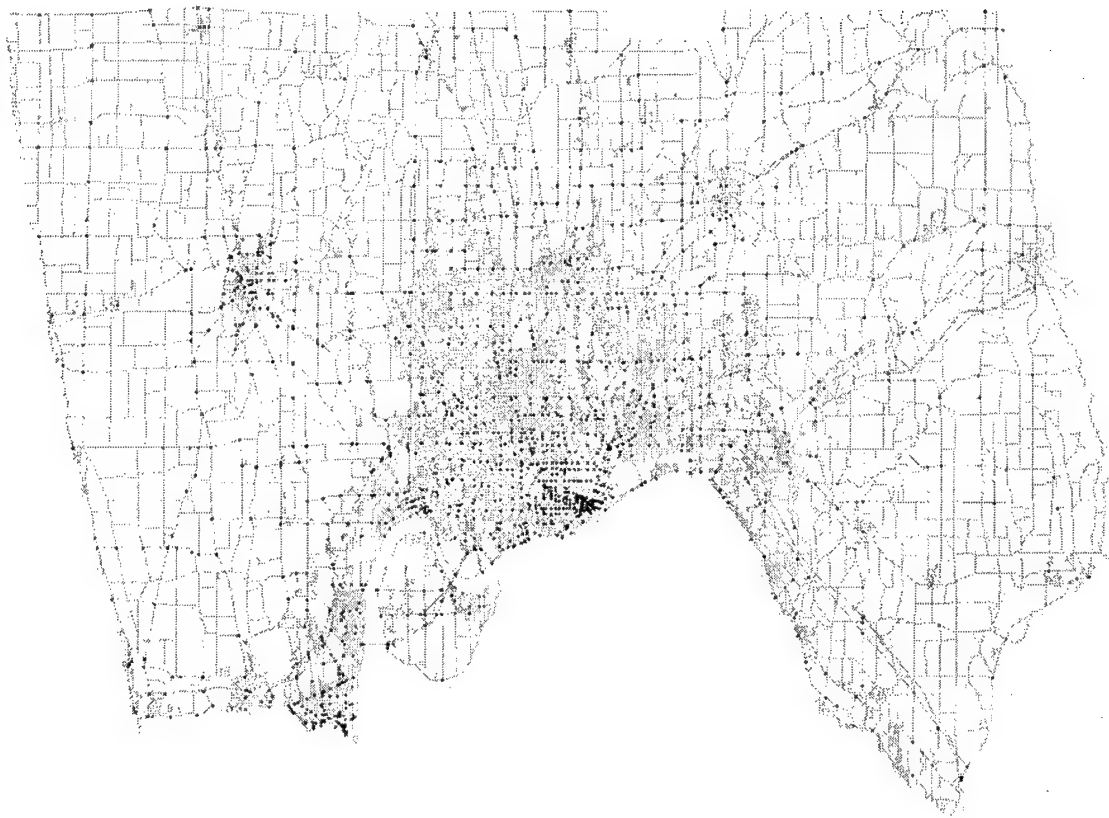
Figure 3. Tailpipe Emission Data Collected and Model Fits
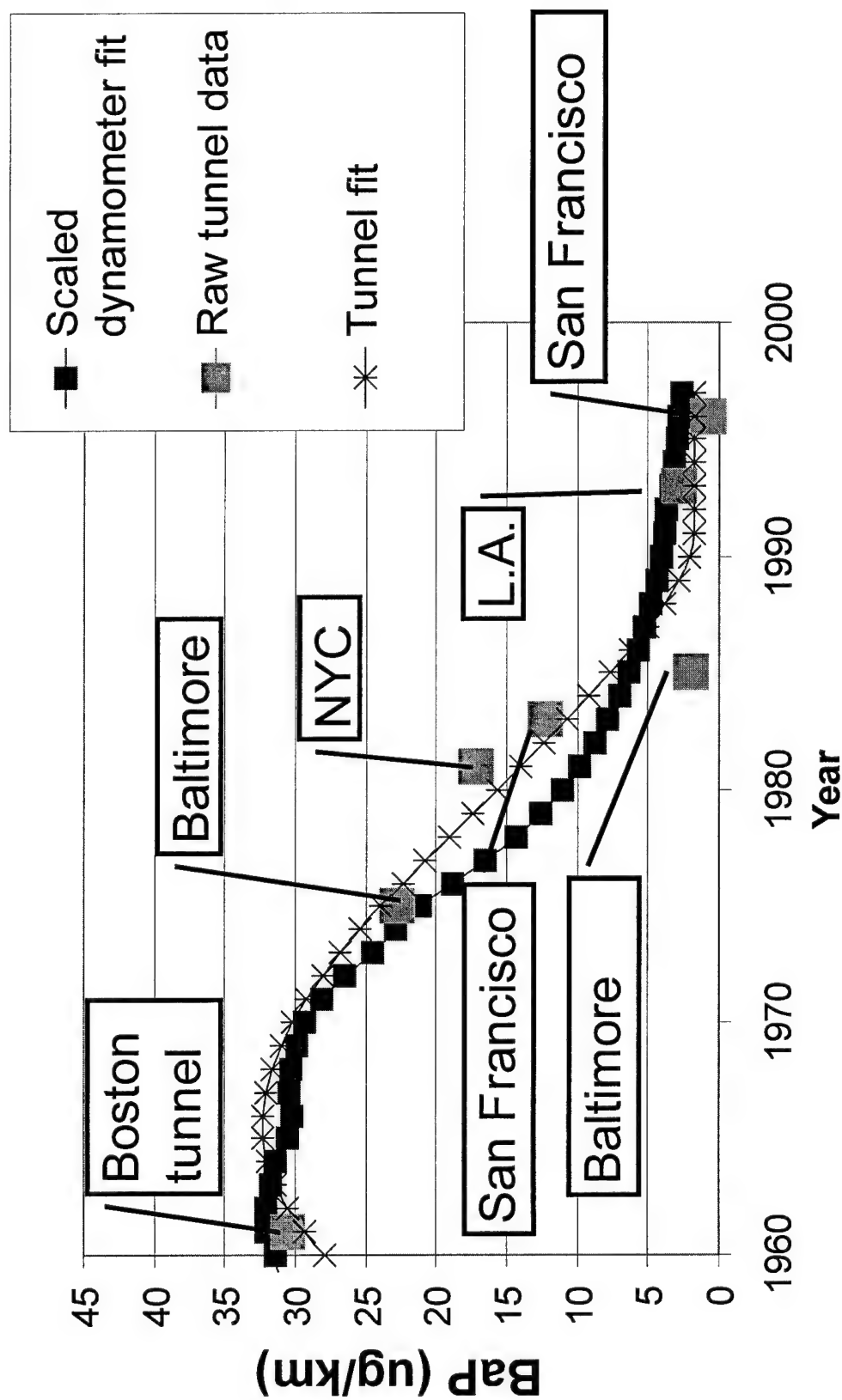
Table 1-1. Characteristics of Study Sample by Case-Control Status, in menarche analyses

Characteristics*	Pre-menopausal		Post-menopausal	
	Cases	Controls	Cases	Controls
N	187	347	52	76
Race (% of whites)	96.3	96.3	92.3	90.8
Age(years)	43.8 (4.0)*	43.0 (3.9)*	49.4 (3.0)*	47.3 (3.0)*
Education (years)	13.8 (2.0)	14.2 (2.2)	14.6 (2.4)	14.0 (2.4)
Body Mass Index (1 year prior to interview)	27.0 (7.0)	27.4 (6.8)	28.3 (5.3)	28.5 (7.1)
Percentage of women without child	17.6	16.1	26.9	22.4
Age at menarche (years)	12.6 (1.6)	12.7 (1.6)	12.9 (1.5)	13.1 (1.8)
Age at menopause	/	/	45.0 (5.1)*	42.1 (6.5)*
Number of births	1.8 (1.2)	2.0 (1.2)	1.6 (1.3)	1.9 (1.4)
First-degree relative with breast cancer (% of Yes)	22.5*	9.8*	17.3	13.2
Previous benign breast disease (% of Yes)	33.2*	19.6*	30.8	22.4
PAH exposure, run5	8.7 (1.0)*	8.4 (1.1)*	8.2 (0.9)	8.5 (1.1)
PAH exposure, run6	8.5 (0.9)*	8.3 (1.0)*	8.0 (0.9)	8.2 (1.0)
PAH exposure, run7	9.8 (0.9)*	9.6 (1.1)*	9.5 (1.0)	9.6 (1.0)
PAH exposure, run8	9.3 (0.9)*	9.1 (1.1)*	9.0 (1.0)	9.1 (1.0)

Distance to closest road with traffic count (meters)	172.9 (213.3)* (median=116.5)	232.2 (458.2)* (median=135.5)	164.4 (294.8) (median=88.9)	197.5 (446.5) (median=102.8)
Number of years stay in this address	16.7 (6.9)	17.3 (6.2)	14.6 (6.4)	15.7 (8.3)
AADT in 1990 of the closest road	12515.4 (15962.1)	11922.1 (15398.8)	11819.9 (16533.3)	12250.8 (16088.1)
Year of appointment (% in 2000)	27.8*	0.9*	40.4*	5.3*
Year exposed (% in 1960)	4.3*	6.1*	28.8	23.7
Income (% of 30,000 or under)	23.8*	13.4*	15.9	17.6
	(n=168)	(n=337)	(n=44)	(n=74)
Smoking (packyear)	7.6 (11.4)*	5.1 (9.0)*	8.6 (15.5)	12.0 (15.8)
	(n=187)	(n=346)	(n=52)	(n=76)
Never Smoker (%)	47.6	53.5	52.9	43.4
	(n=187)	(n=346)	(n=51)	(n=76)

* Values shown are mean (SD) except for race which are percentages of whites, percentage of first-degree relative with breast cancer, and percentage of people with previous benign breast disease, among either the cases or controls. Two-sided comparisons of means between the cases and controls were computed by T-test; comparisons of categories were with the chi-square test. Those with * are significantly different, p<=0.05.

Table 1-2. Characteristics of Study Sample by Case-Control Status, in first birth analyses

Characteristics*	Pre-menopausal		Post-menopausal	
	Cases	Controls	Cases	Controls
N	181	371	221	308
Race (% of whites)	95.6	96.0	93.7	94.2
Age(years)	44.5 (4.7)	44.2 (4.6)	57.0 (6.0)*	54.5 (6.1)*
Education (years)	13.9 (2.0)	14.1 (2.1)	14.1 (2.4)	13.7 (2.3)
Body Mass Index (1 year prior to interview)	27.2 (7.2)	27.3 (6.6)	28.8 (5.6)	28.7 (6.4)
Percentage of women without child	/	/	/	/
Age at menarche (years)	12.5 (1.4)	12.7 (1.6)	12.4 (1.6)	12.5 (1.7)
Age at menopause	/	/	48.5 (4.7)*	46.9 (5.9)*
Number of births	2.2 (0.9)	2.3 (0.9)	2.4 (1.0)*	2.6 (1.2)*
First-degree relative with breast cancer (% of Yes)	21.0*	8.9*	20.4*	10.1*
Previous benign breast disease (% of Yes)	35.4*	21.8*	38.5*	25.6*
PAH exposure, run5	7.4 (1.4)	7.5 (1.3)	8.5 (1.1)	8.4 (1.2)
PAH exposure, run6	7.2 (1.3)	7.3 (1.2)	8.3 (1.0)	8.2 (1.1)
PAH exposure, run7	8.5 (1.3)	8.6 (1.3)	9.6 (1.0)	9.6 (1.1)
PAH exposure, run8	8.1 (1.3)	8.2 (1.3)	9.2 (1.0)	9.1 (1.1)

Distance to closest road with traffic count (meters)	199.8 (233.6) (median=135.5)	244.3 (431.4) (median=132.0)	167.7 (193.0) (median=119.7)	183.9 (310.8) (median=123.2)
Number of years stay in this address	9.6 (7.6)	10.6 (8.0)	11.6 (12.3)	11.0 (10.7)
AADT in 1990 of the closest road	11671.8 (13271.5)	12140.7 (13150.2)	12254.7 (9694.3)	12518.6 (15251.2)
Year of appointment (% in 2000)	27.1*	0.8*	33.9*	2.6*
Year exposed (% in 1970)	2.2	2.2	5.0	6.2
Income (% of 30,000 or under)	22.2* (n=162)	11.3* (n=362)	20.4 (n=191)	22.5 (n=293)
Smoking (packyear)	7.7 (11.4)* (n=181)	5.4 (9.7)* (n=370)	10.2 (13.8) (n=221)	12.6 (17.4) (n=308)
Never Smoker (%)	45.3 (n=181)	53.9 (n=371)	38.2 (n=220)	38.6 (n=308)

* Values shown are mean (SD) except for race which are percentages of whites, percentage of first-degree relative with breast cancer, and percentage of people with previous benign breast disease, among either the cases or controls. Two-sided comparisons of means between the cases and controls were computed by T-test; comparisons of categories were with the chi-square test. Those with * are significantly different, $p \leq 0.05$.

Table 2-1. Correlations among the variables, in Menarche analyses, pre-menopausal women

	Yrs at Appt.	Dist_closed	Yrs at Addr. (n=532)	AADT90	Year at Exp.	Age	Edu. (year)	BMI	Age at Menar.	Parity	BaP (R5)	BaP(R6)	BaP(R7)	BaP(R8)	Income	Smoking (Packyear)
Yrs at Appt.	1															
Dist_closeRd	-0.02	1														
Yrs at Addr. (n=532)	-0.06	0.05	1													
AADT90	0.01	-0.05	0.06	1												
Year at Exp.	0.16*	0.01	0.02	0.03	1											
Age	0.16*	-0.01	-0.07	-0.04	-0.87*	1										
Edu. (year)	-0.04	-0.02	0.04	-0.01	-0.08	0.06	1									
BMI	-0.05	-0.01	-0.11*	-0.03	-0.11*	0.03	-0.17*	1								
Age at Menar.	-0.00	0.01	-0.06	-0.02	0.29*	0.09*	-0.02	-0.18*	1							
Parity	-0.06	-0.00	-0.10*	-0.02	-0.07	0.08	-0.09*	-0.04	0.08	1						
BaP (R5)	0.02	-0.37*	0.04	0.29*	-0.03	0.02	-0.02	-0.00	-0.03	-0.11*	1					
BaP(R6)	0.02	-0.39*	0.05	0.30*	-0.02	0.01	-0.02	-0.01	-0.03	-0.12*	0.99*	1				
BaP(R7)	0.01	-0.48*	0.00	0.34*	-0.06	0.04	-0.03	-0.02	-0.04	-0.11*	0.83*	0.86*	1			
BaP(R8)	0.02	-0.48*	0.00	0.37*	-0.05	0.03	-0.02	-0.03	-0.04	-0.11*	0.80*	0.83*	0.99*	1		
Income	-0.00	-0.07	0.01	0.01	-0.09*	0.09*	0.34*	-0.20*	-0.00	0.09*	0.03	0.03	0.04	0.05	1	
Smoking (Packyear)	0.05	-0.05	-0.11*	0.11*	0.03	-0.03	-0.23*	0.10*	-0.03	0.03	0.08	0.07	0.08	0.07	-0.08	1

Table 2-2. Correlations among the variables, in Menarche analyses, post-menopausal women

	Yrs at Appt.	Dist_close Rd	Yrs at Addr. (n=127)	AADT90	Year at Exp.	Age	Edu. (year)	BMI	Age at Menar.	Age at Menop.	Parity	BaP (R5)	BaP(R6)	BaP(R7)	BaP(R8)	Income	Smoking (Packyear)
Yrs at Appt.	1																
Dist_close Rd	-0.03	1															
Yrs at Addr. (n=127)	-0.06	0.09	1														
AADT90	0.04	0.01	0.06	1													
Year at Exp.	0.01	0.07	-0.10	-0.09	1												
Age	0.39*	-0.09	0.06	0.10	-0.77*	1											
Edu. (year)	0.14	-0.04	0.06	0.04	0.01	0.05	1										
BMI	-0.06	-0.03	-0.01	-0.10	0.04	-0.10	-0.12	1									
Age at Menar.	-0.04	-0.05	0.01	-0.00	0.23*	-0.27*	0.05	-0.09	1								
Age at Menop.	0.32*	-0.15	-0.14	0.06	-0.28*	-0.44*	0.07	-0.13	0.07	1							
Parity	-0.06	-0.01	-0.20*	0.00	-0.05	0.06	-0.15	0.02	-0.02	0.13	1						
BaP (R5)	-0.15	-0.34*	-0.01	0.33*	0.10	-0.13	0.11	0.03	0.03	0.00	0.14	1					
BaP(R6)	-0.13	-0.35*	-0.02	0.32*	0.15	-0.16	0.14	0.02	0.05	-0.02	0.11	0.98*	1				
BaP(R7)	-0.09	-0.45*	-0.01	0.28*	0.14	-0.14	0.12	0.01	0.05	-0.06	-0.00	0.83*	0.87*	1			
BaP(R8)	-0.09	-0.49*	-0.04	0.29*	0.14	-0.14	0.11	0.00	0.05	-0.07	0.01	0.82*	0.86*	0.99*	1		
Income	-0.01	0.01	0.02	-0.10	-0.07	0.04	0.38*	-0.16	-0.01	0.23*	0.01	-0.16	-0.15	-0.07	-0.09	1	
Smoking (Packyear)	-0.14	0.17	-0.18*	-0.05	-0.08	0.02	-0.14	-0.10	-0.02	0.08	-0.02	-0.01	-0.02	0.01	0.02	-0.07	1

Table 2-3. Correlations among the variables, in First-birth analyses, pre-menopausal women

	Yrs at Appt.	Dist_closeRd	Yrs at Addr. (n=550)	AADT90	Year at Exp.	Age	Edu. (year)	BMI	Age at Menar.	Parity	BaP (R5)	BaP(R6)	BaP(R7)	BaP(R8)	Income	Smoking (Packyear)
Yrs at Appt.	1															
Dist_closeRd	0.01	1														
Yrs at Addr. (n=550)	-0.02	0.22*	1													
AADT90	-0.07	-0.04	-0.06	1												
Year at Exp.	0.13*	0.04	-0.04	0.01	1											
Age	0.08*	0.04	0.09*	-0.08	-0.70*	1										
Edu. (year)	-0.01	-0.03	-0.03	-0.01	0.20*	0.04	1									
BMI	-0.04	-0.03	-0.07	-0.05	-0.14*	0.08	-0.13*	1								
Age at Menar.	-0.02	0.04	0.03	-0.02	0.00	0.00	0.00	-0.17*	1							
Parity	-0.05	-0.03	-0.08	0.03	-0.27*	0.08	-0.06	0.01	-0.02	1						
BaP (R5)	-0.12*	-0.32*	-0.08	0.17*	-0.61*	0.34*	-0.09*	0.09*	-0.04	0.18*	1					
BaP(R6)	-0.12*	-0.33*	-0.09*	0.19*	-0.65*	0.37*	-0.12*	0.09*	-0.04	0.19*	0.99*	1				
BaP(R7)	-0.11*	-0.42*	-0.14*	0.25*	-0.65*	0.35*	-0.14*	0.11*	-0.04	0.16*	0.87*	0.91*	1			
BaP(R8)	-0.11*	-0.44*	-0.14*	0.27*	-0.65*	0.35*	-0.13*	0.11*	-0.04	0.16*	0.86*	0.90*	0.99*	1		
Income	0.03	0.03	-0.07	0.05	0.13*	0.05	0.41*	-0.17*	0.00	-0.03	-0.05	-0.06	-0.08	-0.08	1	
Smoking (Packyear)	0.06	-0.04	-0.05	0.04	-0.12*	-0.02	-0.19*	0.08*	-0.03	-0.01	0.13*	0.14*	0.12*	0.12*	-0.07	1

Table 2-4. Correlations among the variables, in First-birth analyses, post-menopausal women

	Yrs at Appt.	Dist_close Rd	Yrs at Addr. (n=528)	AADT90	Year at Exp.	Age	Edu. (year)	BMI	Age at Menar.	Age at Menop.	Parity	BaP (R5)	BaP(R6)	BaP(R7)	BaP(R8)	Income	Smoking (Packyear)
Yrs at Appt.	1																
Dist_close Rd	-0.00	1															
Yrs at Addr. (n=528)	0.07	0.14*	1														
AADT90	0.01	0.01	0.05	1													
Year at Exp.	0.09*	0.09*	0.11*	0.02	1												
Age	0.18*	-0.04	0.15*	0.03	-0.59*	1											
Edu. (year)	0.08	0.06	-0.10*	0.04	0.22*	-0.02	1										
BMI	0.05	0.02	0.10*	-0.08	0.05	-0.04	-0.02	1									
Age at Menar.	-0.02	0.00	0.07	0.03	-0.03	0.10*	-0.05	-0.16*	1								
Age at Menop.	0.07	-0.05	0.09*	-0.00	-0.17*	0.38*	0.10*	0.00	0.02	1							
Parity	-0.12*	-0.08	-0.20*	-0.06	-0.37*	0.03	-0.00	-0.06	-0.01	0.13*	1						
BaP (R5)	-0.04	-0.31*	-0.21*	0.22*	-0.23*	0.11*	-0.01	-0.03	0.03	0.01	0.08	1					
BaP(R6)	-0.04	-0.32*	-0.21*	0.26*	-0.25*	0.12*	-0.02	-0.04	0.03	0.02	0.08	0.98*	1				
BaP(R7)	-0.04	-0.43*	-0.18*	0.33*	-0.27*	0.11*	-0.01	-0.06	0.02	0.02	0.12*	0.82*	0.86*	1			
BaP(R8)	-0.04	-0.45*	-0.17*	0.35*	-0.26*	0.12*	-0.02	-0.07	0.03	0.02	0.11*	0.79*	0.84*	0.99*	1		
Income	0.05	0.03	-0.10*	0.02	0.23*	-0.23*	0.42*	-0.07	0.00	0.00	-0.04	-0.01	-0.01	-0.04	-0.05	1	
Smoking (Packyear)	-0.04	0.04	-0.03	0.02	-0.15*	0.03	0.13*	0.03	-0.08	-0.09*	0.00	0.07	0.07	0.04	0.04	-0.11*	1

67

Table 3. Traffic PAH Exposure in Different Time Periods and Risk of Breast Cancer

Quartiles of PAH Exposure	Cases	Controls	Crude OR (CI)	Adjusted OR (CI)*	P for trend
At Menarche					
Pre-menopausal					
1: Lowest thru 7.65	29	86	1.00	1.00	
2: 7.65 thru 8.36	49	87	1.67 (0.97-2.89)	1.91 (0.86-4.26)	
3: 8.36 thru 8.84	52	89	1.73(1.01-2.98)*	2.05 (0.89-4.73)	
4: 8.84 thru Highest	57	85	1.99(1.16-3.41)*	2.14 (0.94-4.87)	0.03*
Post-menopausal					
1: Lowest thru 7.75	16	19	1.00	1.00	
2: 7.75 thru 8.41	19	19	1.19 (0.47-2.98)	1.03 (0.24-4.40)	
3: 8.41 thru 8.72	6	19	0.38 (0.12-1.17)	0.41 (0.08-2.18)	
4: 8.72 thru Highest	11	19	0.69 (0.25-1.86)	0.46 (0.08-2.79)	0.41
At First Birth					
Pre-menopausal					
1: Lowest thru 6.42	45	92	1.00	1.00	
2: 6.42 thru 7.41	56	92	1.24 (0.76-2.03)	1.96 (0.89-4.32)	
3: 7.41 thru 8.16	40	94	0.87 (0.52-1.45)	0.86 (0.36-2.05)	
4: 8.16 thru Highest	40	93	0.88 (0.53-1.47)	1.36 (0.50-3.72)	0.69
Post-menopausal					
1: Lowest thru 7.57	39	77	1.00	1.00	
2: 7.57 thru 8.35	67	77	1.72(1.04-2.85)*	2.09 (0.93-4.67)	
3: 8.35 thru 8.76	53	78	1.34 (0.80-2.26)	1.19 (0.51-2.77)	
4: 8.76 thru Highest	62	76	1.61 (0.97-2.68)	2.58(1.14-5.82)*	0.19

* Odds ratios and 95% confidence intervals adjusted for age, education, race, BMI (1 yr before interview), age at first birth, age at menarche, age at menopause (for post-menopausal women only), number of births, first-degree relative with breast cancer, previous benign breast disease and year at interview.

Table 4. Traffic PAH Exposure in Different Time Periods and Risk of Breast Cancer (restrict to Distance to closest Rd with traffic count within 250 meters)

Quartiles of PAH Exposure	Cases	Controls	Crude OR (CI)	Adjusted OR (CI)*	P for trend
At Menarche					
Pre-menopausal					
1:	14	54	1.00	1.00	
2:	39	68	2.21(1.09-4.49)*	2.83 (0.94-8.52)	
3:	44	77	2.20(1.10-4.42)*	2.96 (0.98-8.96)	
4:	53	71	2.88(1.45-5.72)*	3.90(1.30-11.69)*	0.02*
Post-menopausal					
1:	11	13	1.00	1.00	
2:	15	16	1.11 (0.38-3.22)	0.92 (0.15-5.54)	
3:	6	16	0.44 (0.13-1.52)	0.68 (0.09-5.19)	
4:	10	18	0.66 (0.22-2.00)	0.50 (0.07-3.71)	0.88
At First Birth					
Pre-menopausal					
1:	24	45	1.00	1.00	
2:	46	67	1.29 (0.69-2.40)	2.21 (0.82-5.90)	
3:	35	76	0.86 (0.46-1.63)	1.03 (0.36-2.90)	
4:	33	79	0.78 (0.41-1.49)	1.33 (0.40-4.45)	0.78
Post-menopausal					
1:	24	50	1.00	1.00	
2:	55	61	1.88(1.02-3.45)*	2.03 (0.77-5.31)	
3:	41	70	1.22 (0.66-2.27)	0.91 (0.32-2.56)	
4:	52	66	1.64 (0.89-3.01)	2.63(1.01-6.88)*	0.25

* Odds ratios and 95% confidence intervals adjusted for age, education, race, BMI (1 yr before interview), age at first birth, age at menarche, age at menopause (for post-menopausal women only), number of births, first-degree relative with breast cancer, previous benign breast disease and year at interview.

Table 5. Traffic PAH Exposure in Different Time Periods and Risk of Breast Cancer (adjusting for income instead of education)

Quartiles of PAH Exposure	Cases	Controls	Crude OR (CI)	Adjusted OR (CI)*	P for trend
At Menarche					
Pre-menopausal					
1:	26	83	1.00	1.00	
2:	42	83	1.62 (0.91-2.87)	1.80 (0.79-4.13)	
3:	47	87	1.73 (0.98-3.04)	2.07 (0.89-4.83)	
4:	53	84	2.01(1.15-3.52)*	2.07 (0.90-4.78)	0.03*
Post-menopausal					
1:	11	19	1.00	1.00	
2:	16	19	1.46 (0.54-3.94)	1.26 (0.27-5.83)	
3:	6	18	0.58 (0.18-1.88)	0.50 (0.09-2.69)	
4:	11	18	1.06 (0.37-3.03)	0.62 (0.10-4.06)	0.58
At First Birth					
Pre-menopausal					
1:	37	90	1.00	1.00	
2:	51	91	1.36 (0.82-2.28)	1.94 (0.87-4.32)	
3:	38	89	1.04 (0.61-1.78)	0.86 (0.35-2.09)	
4:	36	92	0.95 (0.55-1.64)	1.39 (0.50-3.84)	0.87
Post-menopausal					
1:	33	74	1.00	1.00	
2:	55	74	1.67 (0.97-2.86)	1.97 (0.82-4.78)	
3:	46	73	1.41 (0.81-2.45)	1.21 (0.48-3.00)	
4:	57	72	1.78(1.04-3.04)*	2.87(1.19-6.91)*	0.11

* Odds ratios and 95% confidence intervals adjusted for age, income, race, BMI (1 yr before interview), age at first birth, age at menarche, age at menopause (for post-menopausal women only), number of births, first-degree relative with breast cancer, previous benign breast disease and year at interview.

Table 6. Traffic PAH Exposure in Different Time Periods and Risk of Breast Cancer (adjusting for both income and education)

Quartiles of PAH Exposure	Cases	Controls	Crude OR (CI)	Adjusted OR (CI)*	P for trend
At Menarche					
Pre-menopausal					
1:	26	83	1.00	1.00	
2:	42	83	1.62 (0.91-2.87)	1.80 (0.78-4.12)	
3:	47	87	1.73 (0.98-3.04)	2.07 (0.89-4.83)	
4:	53	84	2.01(1.15-3.52)*	2.07 (0.90-4.77)	0.03*
Post-menopausal					
1:	11	19	1.00	1.00	
2:	16	19	1.46 (0.54-3.94)	1.26 (0.27-5.92)	
3:	6	18	0.58 (0.18-1.88)	0.53 (0.10-2.93)	
4:	11	18	1.06 (0.37-3.03)	0.61 (0.09-4.02)	0.49
At First Birth					
Pre-menopausal					
1:	37	90	1.00	1.00	
2:	51	91	1.36 (0.82-2.28)	2.02 (0.90-4.52)	
3:	38	89	1.04 (0.61-1.78)	0.86 (0.35-2.10)	
4:	36	92	0.95 (0.55-1.64)	1.41 (0.51-3.90)	0.88
Post-menopausal					
1:	33	74	1.00	1.00	
2:	55	74	1.67 (0.97-2.86)	1.98 (0.82-4.78)	
3:	46	73	1.41 (0.81-2.45)	1.21 (0.48-3.00)	
4:	57	72	1.78(1.04-3.04)*	2.87(1.19-6.92)*	0.11

74

*   Odds ratios and 95% confidence intervals adjusted for age, education, income, race, BMI (1 yr before interview), age at first birth, age at menarche, age at menopause (for post-menopausal women only), number of births, first-degree relative with breast cancer, previous benign breast disease and year at interview.

Table 7-1. Traffic PAH Exposure in Different Time Periods and Risk of Breast Cancer (in Situ cancer only)

Quartiles of PAH Exposure	Cases	Controls	Crude OR (CI)	Adjusted OR (CI)*	P for trend
At Menarche					
Pre-menopausal					
1:	4	86	1.00	1.00	
2:	10	87	2.47 (0.75-8.18)	7.21(0.73-71.12)	
3:	6	89	1.45 (0.40-5.32)	1.05(0.08-13.88)	
4:	10	85	2.53 (0.76-8.38)	10.57(0.95-117.41)	0.18
Post-menopausal					
1:	3	19	1.00	1.00	
2:	4	19	1.33 (0.26-6.78)	0.03(0.00-)	
3:	0	19	0.00(0.00-2.8E+28)	0.00(0.00-)	
4:	1	19	0.33 (0.03-3.50)	3.98(0.00-)	1.00
At First Birth					
Pre-menopausal					
1:	5	92	1.00	1.00	
2:	13	92	2.60 (0.89-7.59)	4.42(0.79-24.64)	
3:	8	94	1.57 (0.49-4.96)	0.94 (0.14-6.33)	
4:	3	93	0.59 (0.14-2.56)	1.46(0.12-17.52)	0.48
Post-menopausal					
1:	4	77	1.00	1.00	
2:	12	77	3.00 (0.93-9.71)	3.27(0.50-21.34)	
3:	4	78	0.99 (0.24-4.09)	0.42 (0.05-3.94)	
4:	8	76	2.03 (0.59-7.01)	8.99(1.17-69.04)*	0.02*

* Odds ratios and 95% confidence intervals adjusted for age, education, race, BMI (1 yr before interview), age at first birth, age at menarche, age at menopause (for post-menopausal women only), number of births, first-degree relative with breast cancer, previous benign breast disease and year at interview.

Table 7-2. Traffic PAH Exposure in Different Time Periods and Risk of Breast Cancer (in Situ cancer only), model 2, with fewer covariates

Quartiles of PAH Exposure	Cases	Controls	Crude OR (CI)	Adjusted OR (CI)*	P for trend
At Menarche					
Pre-menopausal					
1:	4	86	1.00	1.00	
2:	10	87	2.47 (0.75-8.18)	3.94(0.67-23.13)	
3:	6	89	1.45 (0.40-5.32)	1.11 (0.13-9.20)	
4:	10	85	2.53 (0.76-8.38)	4.11(0.70-24.16)	0.30
Post-menopausal					
1:	3	19	1.00	1.00	
2:	4	19	1.33 (0.26-6.78)	0.68(0.03-16.50)	
3:	0	19	0.00(0.00-2.8E+28)	0.00(0.00-1.25E+42)	
4:	1	19	0.33 (0.03-3.50)	0.03 (0.00-1.26)	0.13
At First Birth					
Pre-menopausal					
1:	5	92	1.00	1.00	
2:	13	92	2.60 (0.89-7.59)	5.07(1.04-24.62)*	
3:	8	94	1.57 (0.49-4.96)	0.94 (0.17-5.21)	
4:	3	93	0.59 (0.14-2.56)	1.76(0.29-10.77)	0.65
Post-menopausal					
1:	4	77	1.00	1.00	
2:	12	77	3.00 (0.93-9.71)	2.75(0.60-12.64)	
3:	4	78	0.99 (0.24-4.09)	0.29 (0.04-1.96)	
4:	8	76	2.03 (0.59-7.01)	4.42(0.87-22.36)	0.11

*   Odds ratios and 95% confidence intervals adjusted for year at interview only, except that post-menopausal at first birth analyses (additionally adjusted for BMI (1 yr before interview)); and post-menopausal 20 year prior and 10 year prior analyses (additionally adjusted for age, age at menopause, number of births, and previous benign breast disease).

Table 8. Traffic PAH Exposure in Different Time Periods and Risk of Breast Cancer (non-inSitu cancer only)

Quartiles of PAH Exposure	Cases	Controls	Crude OR (CI)	Adjusted OR (CI)*	P for trend
At Menarche					
Pre-menopausal					
1:	24	86	1.00	1.00	
2:	39	87	1.61 (0.89-2.90)	1.75 (0.76-4.03)	
3:	44	89	1.77 (0.99-3.16)	2.02 (0.86-4.74)	
4:	46	85	1.94(1.09-3.46)*	1.74 (0.74-4.12)	0.07
Post-menopausal					
1:	13	19	1.00	1.00	
2:	14	19	1.08 (0.40-2.89)	0.91 (0.19-4.23)	
3:	5	19	0.39 (0.12-1.29)	0.30 (0.05-1.81)	
4:	10	19	0.77 (0.27-2.18)	0.49 (0.08-3.22)	0.38
At First Birth					
Pre-menopausal					
1:	39	92	1.00	1.00	
2:	41	92	1.05 (0.62-1.78)	1.71 (0.74-3.96)	
3:	32	94	0.80 (0.46-1.39)	0.77 (0.31-1.94)	
4:	36	93	0.91 (0.53-1.56)	1.43 (0.51-4.06)	0.81
Post-menopausal					
1:	34	77	1.00	1.00	
2:	55	77	1.62 (0.95-2.75)	2.18 (0.95-5.00)	
3:	48	78	1.39 (0.81-2.39)	1.09 (0.46-2.60)	
4:	52	76	1.55 (0.91-2.65)	2.42(1.03-5.67)*	0.51

*   Odds ratios and 95% confidence intervals adjusted for age, education, race, BMI (1 yr before interview), age at first birth, age at menarche, age at menopause (for post-menopausal women only), number of births, first-degree relative with breast cancer, previous benign breast disease and year at interview.

Table 9. Traffic PAH Exposure in Different Time Periods and Risk of Breast Cancer (non-smoker only)

Quartiles of PAH Exposure	Cases	Controls	Crude OR (CI)	Adjusted OR (CI)*	P for trend
At Menarche					
Pre-menopausal					
1:	11	54	1.00	1.00	
2:	25	45	2.73(1.21-6.14)*	4.66(1.19-18.17)*	
3:	26	45	2.84(1.26-6.37)*	6.94(1.76-27.37)*	
4:	27	41	3.23(1.44-7.27)*	6.98(1.76-27.72)*	0.01*
Post-menopausal					
1:	9	6	1.00	1.00	
2:	9	7	0.86 (0.21-3.58)	0.56 (0.04-8.88)	
3:	3	11	0.18(0.04-0.94)*	0.13 (0.01-1.61)	
4:	6	9	0.44 (0.10-1.92)	0.85(0.04-19.36)	0.43
At First Birth					
Pre-menopausal					
1:	20	56	1.00	1.00	
2:	31	47	1.85 (0.93-3.66)	4.06(1.16-14.21)*	
3:	17	48	0.99 (0.47-2.11)	1.38 (0.36-5.37)	
4:	14	49	0.80 (0.37-1.75)	1.49 (0.27-8.13)	0.30
Post-menopausal					
1:	16	41	1.00	1.00	
2:	30	29	2.65(1.23-5.73)*	5.11(1.37-19.07)*	
3:	13	25	1.33 (0.55-3.23)	2.52(0.57-11.20)	
4:	25	24	2.67(1.19-5.97)*	8.53(2.14-34.00)*	0.04*

*   Odds ratios and 95% confidence intervals adjusted for age, education, race, BMI (1 yr before interview), age at first birth, age at menarche, age at menopause (for post-menopausal women only), number of births, first-degree relative with breast cancer, previous benign breast disease and year at interview.

Table 10. Traffic PAH Exposure in Different Time Periods and Risk of Breast Cancer (smoker only)

Quartiles of PAH Exposure	Cases	Controls	Crude OR (CI)	Adjusted OR (CI)*	P for trend
At Menarche					
Pre-menopausal					
1:	**18**	**32**	**1.00**	**1.00**	
2:	**24**	**42**	**1.02 (0.47-2.18)**	**0.85 (0.28-2.59)**	
3:	**26**	**44**	**1.05 (0.49-2.23)**	**0.85 (0.25-2.83)**	
4:	**30**	**43**	**1.24 (0.59-2.61)**	**0.75 (0.24-2.35)**	**0.99**
Post-menopausal					
1:	**7**	**13**	**1.00**	**1.00**	
2:	**9**	**12**	**1.39 (0.40-4.92)**	**0.26(0.00-15.27)**	
3:	**3**	**8**	**0.70 (0.14-3.50)**	**0.31(0.00-37.13)**	
4:	**5**	**10**	**0.93 (0.23-3.82)**	**0.05 (0.00-9.25)**	**0.37**
At First Birth					
Pre-menopausal					
1:	**25**	**36**	**1.00**	**1.00**	
2:	**25**	**45**	**0.80 (0.40-1.62)**	**0.98 (0.31-3.09)**	
3:	**23**	**46**	**0.72 (0.35-1.47)**	**0.51 (0.14-1.80)**	
4:	**26**	**44**	**0.85 (0.42-1.72)**	**0.95 (0.25-3.63)**	**0.58**
Post-menopausal					
1:	**22**	**36**	**1.00**	**1.00**	
2:	**37**	**48**	**1.26 (0.64-2.50)**	**1.14 (0.38-3.40)**	
3:	**40**	**53**	**1.24 (0.63-2.42)**	**0.75 (0.25-2.27)**	
4:	**37**	**52**	**1.16 (0.59-2.29)**	**1.24 (0.42-3.68)**	**0.94**

84

* Odds ratios and 95% confidence intervals adjusted for age, education, race, BMI (1 yr before interview), age at first birth, age at menarche, age at menopause (for post-menopausal women only), number of births, first-degree relative with breast cancer, previous benign breast disease and year at interview.

Table 11. Traffic PAH Exposure in Different Time Periods and Risk of Breast Cancer (adjust for packyear of smoking)

Quartiles of PAH Exposure	Cases	Controls	Crude OR (CI)	Adjusted OR (CI)*	P for trend
At Menarche					
Pre-menopausal					
1:	29	86	1.00	1.00	
2:	49	87	1.67 (0.97-2.89)	1.90 (0.85-4.27)	
3:	52	89	1.73(1.01-2.98)*	2.01 (0.87-4.64)	
4:	57	84	2.01(1.17-3.45)*	2.08 (0.91-4.74)	0.04*
Post-menopausal					
1:	16	19	1.00	1.00	
2:	19	19	1.19 (0.47-2.98)	1.04 (0.24-4.47)	
3:	6	19	0.38 (0.12-1.17)	0.40 (0.07-2.16)	
4:	11	19	0.69 (0.25-1.86)	0.47 (0.08-2.85)	0.41
At First Birth					
Pre-menopausal					
1:	45	92	1.00	1.00	
2:	56	91	1.26 (0.77-2.05)	1.96 (0.88-4.35)	
3:	40	94	0.87 (0.52-1.45)	0.85 (0.35-2.02)	
4:	40	93	0.88 (0.53-1.47)	1.29 (0.47-3.55)	0.77
Post-menopausal					
1:	39	77	1.00	1.00	
2:	67	77	1.72(1.04-2.85)*	2.19 (0.97-4.93)	
3:	53	78	1.34 (0.80-2.26)	1.31 (0.56-3.07)	
4:	62	76	1.61 (0.97-2.68)	2.78(1.22-6.32)*	0.15

86

*   Odds ratios and 95% confidence intervals adjusted for age, education, race, BMI (1 yr before interview), age at first birth, age at menarche, age at menopause (for post-menopausal women only), number of births, first-degree relative with breast cancer, previous benign breast disease, year at interview and pack year of smoking.

Table 12. Traffic PAH Exposure in Different Time Periods and Risk of Breast Cancer (ER- only)

Quartiles of PAH Exposure	Cases	Controls	Crude OR (CI)	Adjusted OR (CI)*	P for trend
		At Menarche			
1:	11	105	1.00	1.00	
2:	14	106	1.26 (0.55-2.90)	1.86 (0.58-5.93)	
3:	14	108	1.24 (0.54-2.85)	1.86 (0.55-6.35)	
4:	17	104	1.56 (0.70-3.49)	1.55 (0.47-5.07)	0.13
		At First Birth			
1:	21	169	1.00	1.00	
2:	29	169	1.38 (0.76-2.52)	1.91 (0.87-4.21)	
3:	18	172	0.84 (0.43-1.64)	0.30(0.11-0.82)*	
4:	17	169	0.81 (0.41-1.59)	1.03 (0.39-2.73)	0.57

* **Odds ratios and 95% confidence intervals adjusted for age, education, race, BMI (1 yr before interview), age at first birth, age at menarche, menopausal status, number of births, first-degree relative with breast cancer, previous benign breast disease and year at interview.**

Table 13. Traffic PAH Exposure in Different Time Periods and Risk of Breast Cancer (ER+ only)

Quartiles of PAH Exposure	Cases	Controls	Crude OR (CI)	Adjusted OR (CI)*	P for trend
		At Menarche			
1:	27	105	1.00	1.00	
2:	35	106	1.28 (0.73-2.27)	1.04 (0.46-2.33)	
3:	30	108	1.08 (0.60-1.94)	0.89 (0.38-2.12)	
4:	35	104	1.31 (0.74-2.32)	0.98 (0.42-2.30)	0.78
		At First Birth			
1:	49	169	1.00	1.00	
2:	70	169	1.43 (0.94-2.18)	1.69 (0.88-3.25)	
3:	54	172	1.08 (0.70-1.68)	0.90 (0.45-1.82)	
4:	60	169	1.22 (0.79-1.89)	1.54 (0.76-3.12)	0.97

* **Odds ratios and 95% confidence intervals adjusted for age, education, race, BMI (1 yr before interview), age at first birth, age at menarche, menopausal status, number of births, first-degree relative with breast cancer, previous benign breast disease and year at interview.**

Table 14. Traffic PAH Exposure in Different Time Periods and Risk of Breast Cancer (PR- only)

Quartiles of PAH Exposure	Cases	Controls	Crude OR (CI)	Adjusted OR (CI)*	P for trend
		At Menarche			
1:	15	105	1.00	1.00	
2:	19	106	1.26 (0.61-2.60)	1.72 (0.62-4.72)	
3:	18	108	1.17 (0.56-2.44)	1.70 (0.59-4.90)	
4:	24	104	1.62 (0.80-3.25)	1.72 (0.62-4.79)	0.10
		At First Birth			
1:	34	169	1.00	1.00	
2:	41	169	1.21 (0.73-1.99)	1.52 (0.76-3.06)	
3:	21	172	0.61 (0.34-1.09)	0.28 (0.12-0.68)*	
4:	31	169	0.91 (0.54-1.55)	1.14 (0.50-2.56)	0.47

*   **Odds ratios and 95% confidence intervals adjusted for age, education, race, BMI (1 yr before interview), age at first birth, age at menarche, menopausal status, number of births, first-degree relative with breast cancer, previous benign breast disease and year at interview.**

Table 15. Traffic PAH Exposure in Different Time Periods and Risk of Breast Cancer (PR+ only)

Quartiles of PAH Exposure	Cases	Controls	Crude OR (CI)	Adjusted OR (CI)*	P for trend
		At Menarche			
1:	23	105	1.00	1.00	
2:	31	106	1.34 (0.73-2.44)	1.05 (0.45-2.46)	
3:	26	108	1.10 (0.59-2.05)	0.82 (0.33-2.04)	
4:	28	104	1.23 (0.67-2.27)	0.79 (0.31-1.98)	0.94
		At First Birth			
1:	36	169	1.00	1.00	
2:	58	169	1.61(1.01-2.57)*	2.15(1.06-4.38)*	
3:	52	172	1.42 (0.88-2.28)	1.18 (0.56-2.50)	
4:	46	169	1.28 (0.79-2.08)	1.81 (0.84-3.91)	0.66

*   Odds ratios and 95% confidence intervals adjusted for age, education, race, BMI (1 yr before interview), age at first birth, age at menarche, menopausal status, number of births, first-degree relative with breast cancer, previous benign breast disease and year at interview.

Table 16. Traffic PAH Exposure in Different Years and Risk of Breast Cancer

Quartiles of PAH Exposure	Cases	Controls	Crude OR (CI)	Adjusted OR (CI)*	P for trend
1960					
Pre-menopausal					
1: <7.56	30	98	1.00	1.00	
2: 7.56~8.20	50	108	1.51 (0.89-2.57)	1.41 (0.64-3.10)	
3: 8.20~8.56	47	76	2.02(1.17-3.49)*	2.41(1.04-5.61)*	
4: >8.56	44	82	1.75(1.01-3.04)*	2.09 (0.91-4.80)	0.04*
Post-menopausal					
1: <7.56	123	236	1.00	1.00	
2: 7.56~8.20	172	231	1.43(1.07-1.92)*	1.21 (0.84-1.75)	
3: 8.20~8.56	130	257	0.97 (0.72-1.31)	0.81 (0.56-1.19)	
4: >8.56	140	254	1.06 (0.78-1.43)	1.01 (0.69-1.48)	0.94
1965					
Pre-menopausal					
1: <7.97	33	108	1.00	1.00	
2: 7.97~8.54	66	112	1.93(1.18-3.16)*	2.11(1.03-4.35)*	
3: 8.54~8.92	60	88	2.23(1.34-3.71)*	1.72 (0.78-3.81)	
4: >8.92	50	84	1.95(1.15-3.29)*	2.33(1.07-5.06)*	0.03*
Post-menopausal					
1: <7.97	131	256	1.00	1.00	
2: 7.97~8.54	169	262	1.26 (0.95-1.68)	1.02 (0.72-1.45)	
3: 8.54~8.92	151	273	1.08 (0.81-1.44)	0.91 (0.64-1.30)	
4: >8.92	146	285	1.00 (0.75-1.34)	0.83 (0.57-1.19)	0.33

Table 16. Traffic PAH Exposure in Different Years and Risk of Breast Cancer (continue)

Quartiles of PAH Exposure	Cases	Controls	Crude OR (CI)	Adjusted OR (CI)*	P for trend
1970					
Pre-menopausal					
1: <8.29	35	111	1.00	1.00	
2: 8.29~8.85	70	114	1.95(1.20-3.16)*	2.16(1.05-4.44)*	
3: 8.85~9.23	43	87	1.57 (0.93-2.66)	1.12 (0.50-2.52)	
4: >9.23	57	86	2.10(1.27-3.49)*	2.04 (0.95-4.40)	0.04*
Post-menopausal					
1: <8.29	151	270	1.00	1.00	
2: 8.29~8.85	180	278	1.16 (0.88-1.52)	0.99 (0.71-1.39)	
3: 8.85~9.23	157	298	0.94 (0.71-1.24)	0.84 (0.59-1.18)	
4: >9.23	149	296	0.90 (0.68-1.19)	0.75 (0.53-1.06)	0.19
1980					
Pre-menopausal					
1: <6.73	43	104	1.00	1.00	
2: 6.73~7.53	48	113	1.03 (0.63-1.68)	1.12 (0.54-2.33)	
3: 7.53~7.90	67	120	1.35 (0.85-2.15)	1.49 (0.72-3.08)	
4: >7.90	48	98	1.19 (0.72-1.94)	1.05 (0.47-2.37)	0.29
Post-menopausal					
1: <6.73	164	305	1.00	1.00	
2: 6.73~7.53	183	299	1.14 (0.87-1.48)	1.05 (0.76-1.45)	
3: 7.53~7.90	176	285	1.15 (0.88-1.50)	1.08 (0.78-1.50)	
4: >7.90	159	313	0.95 (0.72-1.24)	0.84 (0.60-1.18)	0.74

1990

			Pre-menopausal		
1: <4.75	54	135	1.00	1.00	
2: 4.75~5.53	65	160	1.02 (0.66-1.56)	1.18 (0.64-2.18)	
3: 5.53~6.00	75	119	1.58(1.03-2.42)*	1.64 (0.86-3.11)	
4: >6.00	59	109	1.35 (0.87-2.12)	1.71 (0.83-3.56)	0.13
			Post-menopausal		
1: <4.75	155	319	1.00	1.00	
2: 4.75~5.53	191	289	1.36(1.04-1.77)*	1.36 (0.98-1.89)	
3: 5.53~6.00	210	330	1.31(1.01-1.70)*	1.37 (0.99-1.89)	
4: >6.00	161	346	0.96 (0.73-1.25)	0.84 (0.60-1.20)	0.56

* Odds ratios and 95% confidence intervals adjusted for age, education, race, BMI (1 yr before interview), age at first birth, age at menarche, age at menopause (for post-menopausal women only), number of births, first-degree relative with breast cancer, previous benign breast disease and year at interview.

Task 3: To evaluate genetic susceptibility in relation to these exposures and breast cancer risk by examining genetic variability in metabolism by NQO1, GSTM1, GST Pi and CYP1A1.

Blood samples are at Dr. Peter Shields' laboratory (Lombardi Cancer Center, Georgetown University). The genotyping analysis for GSTPi and NQO1 are currently in process. Also, data analyses are still underway, thus the results showed here are preliminary, and only include GSTM1 and CYP1A1.

PAH Genetic Susceptibility and Risk of Breast Cancer
PAHs are ubiquitous and occur in the ambient environment at low levels. Using classic epidemiological research methods to detect a potentially small effect on breast cancer may be difficult. As gene-environmental interactions play a more and more important role in the study of a hypothesized weak association between environmental factors and cancer risk, more studies are needed to detect the role of genetic polymorphisms in breast cancer. Polymorphisms in the enzymes involving PAHs metabolism (bioactivation or detoxification) may determine the individual's susceptibility to the formation of DNA adduct. A number of papers have discussed the possible effect of CYP1A1, GSTM1 and GSTP1 on the risk of breast cancer in relation to PAHs. CYP1A1 is the cytochrome P450 isoenzyme involved in the activation of PAHs. Mutations in the CYP1A1 gene may increase the susceptibility of individuals to DNA damage. Glutathione S-transferase (GST) is involved in the detoxification of PAHs by catalyzing conjugation of glutathione with diol epoxides, thus preventing the formation of the PAH-DNA adduct. Among many classes of GST, the GSTM1 genetic polymorphism is a deletion of the entire gene, with a large proportion of people with the null allele. Lack of GSTM1 enzyme may be associated with a higher breast cancer risk. GSTP1, another class of GST, may also play an important role in the breast cancer etiology. One study shows, GSTM1 null genotype associated with increased PAH-DNA adducts in breast cancer cases. However, the role of these polymorphisms is still not clear, and more epidemiological studies are needed to test or confirm the effects of these polymorphisms.

Laboratory Methods
Blood clots for participants were removed from the freezer and shipped on dry ice to Dr. Peter Shields' laboratory (the Lombardi Cancer Center, Georgetown Medical Center) for DNA extraction and genotyping analysis. For those who do not have blood samples, we used urine and saliva samples instead. The genes that are examined in the study include GST M1-1, GST P1-1 and CYP1A1. To control the quality of these analyses, a positive and a negative control were included for every 20 samples. We have included blind duplicates in the samples. Case and control status, as well as other characteristics of the study subjects, are not known by the technicians who perform these tests. These analyses are currently underway.

Results
GSTM1 null genotype was not associated with breast cancer risk (Table 1-1), even after we restricted the analysis to only whites (Table 1-2). CYP1A1 mutant genotype was associated with decreased risk of breast cancer (Table 4-1); the same

tendency and magnitude held even after we restricted the analysis to only whites, but no longer statistically significant (Table 4-2).

In the stratified analyses by GSTM1 genotypes, the originally observed association between PAH exposure and increased breast cancer risk among post-menopausal at first birth analyses remained only in GSTM1 null genotype (Table 2 and 3). Due to the small sample size, the stratified analyses for CYP1A1 genotypes, particular CYP1A1 mutant genotype, was not able to conduct (Table 5 and 6).

Discussion

Consistent with our hypothesis and previous studies, our data suggested that GSTM1 null genotype is not associated with breast cancer risk; and high PAH exposure was only associated with increased breast cancer risk among post-menopausal women with GSTM1 null genotype. The results of CYP1A1 and breast cancer risk were somewhat unexpected, and these preliminary findings are currently under further investigation.

As for most of studies focusing on gene and environmental interaction, the power of this study is limited in stratified analyses when. Because of missing residential address in earlier life, especially at birth, the power of this study is limited by small sample size, preventing us from examining gene-gene interactions. The metabolism of PAH may be affected by multiple genes, and other compounds may also affect the enzyme activity involving in PAH regulation.

96

Table 1-1. GSTM1 and Risk of Breast Cancer

Genotype	Cases	Controls	Crude OR (CI)	Adjusted OR (CI)*
At Menarche				
Pre-menopausal				
1: wt	78	154	1.00	1.00
2: null	62	158	0.78 (0.52-1.16)	0.85 (0.55-1.29)
Post-menopausal				
1: wt	22	32	1.00	1.00
2: null	20	36	0.81 (0.37-1.75)	0.70 (0.27-1.82)
At First Birth				
Pre-menopausal				
1: wt	76	159	1.00	1.00
2: null	64	173	0.77(0.52-1.15)	0.81 (0.54-1.23)
Post-menopausal				
1: wt	79	131	1.00	1.00
2: null	98	122	1.33 (0.91-1.96)	1.22 (0.80-1.85)

* **Odds ratios and 95% confidence intervals adjusted for age, education, race, BMI (1 yr before interview), age at first birth, age at menarche, age at menopause (for post-menopausal women only), number of births, first-degree relative with breast cancer, and previous benign breast disease.**

Table 1-2. GSTM1 and Risk of Breast Cancer (whites only)

Genotype	Cases	Controls	Crude OR (CI)	Adjusted OR (CI)*
At Menarche				
Pre-menopausal				
1: wt	75	144	1.00	1.00
2: null	62	156	0.76 (0.51-1.15)	0.85 (0.56-1.31)
Post-menopausal				
1: wt	20	28	1.00	1.00
2: null	20	35	0.80 (0.36-1.77)	0.77 (0.28-2.12)
At First Birth				
Pre-menopausal				
1: wt	72	151	1.00	1.00
2: null	64	166	0.81 (0.54-1.21)	0.86 (0.57-1.31)
Post-menopausal				
1: wt	71	120	1.00	1.00
2: null	95	119	1.35 (0.91-2.01)	1.24 (0.81-1.90)

* Odds ratios and 95% confidence intervals adjusted for age, education, race, BMI (1 yr before interview), age at first birth, age at menarche, age at menopause (for post-menopausal women only), number of births, first-degree relative with breast cancer, and previous benign breast disease.

Table 2. Traffic PAH Exposure in Different Time Periods and Risk of Breast Cancer (GSTM1 wt)

Quartiles of PAH Exposure	Cases	Controls	Crude OR (CI)	Adjusted OR (CI)*	P for trend
At Menarche					
Pre-menopausal					
1:	11	36	1.00	1.00	
2:	19	36	1.73 (0.72-4.14)	3.00 (0.83-10.80)	
3:	25	38	2.15 (0.93-5.00)	4.17 (1.08-16.17)*	
4:	23	44	1.71 (0.74-3.97)	2.74 (0.71-10.66)	0.12
Post-menopausal					
1:	3	6	1.00	1.00	
2:	11	8	2.75 (0.52-14.44)	1.39 (0.08-23.70)	
3:	3	9	0.67 (0.10-4.48)	0.23 (0.01-5.74)	
4:	5	9	1.11 (0.19-6.49)	0.26 (0.01-6.78)	0.22
At First Birth					
Pre-menopausal					
1:	17	33	1.00	1.00	
2:	31	39	1.54 (0.73-3.27)	2.42 (0.69-8.53)	
3:	14	42	0.65 (0.28-1.50)	0.68 (0.16-2.85)	
4:	14	45	0.60 (0.26-1.40)	0.94 (0.18-4.80)	0.36
Post-menopausal					
1:	15	26	1.00	1.00	
2:	26	34	1.33 (0.59-3.00)	0.97 (0.28-3.34)	

3:	20	32	1.08 (0.47-2.53)	0.47 (0.12-1.84)
4:	18	39	0.80 (0.34-1.86)	1.05 (0.30-3.70)	0.68

*   Odds ratios and 95% confidence intervals adjusted for age, education, race, BMI (1 yr before interview), age at first birth, age at menarche, age at menopause (for post-menopausal women only), number of births, first-degree relative with breast cancer, previous benign breast disease and year at interview.

Table 3. Traffic PAH Exposure in Different Time Periods and Risk of Breast Cancer (GSTM1 null)

Quartiles of PAH Exposure	Cases	Controls	Crude OR (CI)	Adjusted OR (CI)*	P for trend
At Menarche					
Pre-menopausal					
1:	9	42	1.00	1.00	
2:	23	40	2.68 (1.11-6.50)*	1.98 (0.52-7.56)	
3:	13	42	1.44 (0.56-3.74)	1.06 (0.26-4.32)	
4:	17	34	2.33 (0.92-5.89)	2.44 (0.63-9.46)	0.18
Post-menopausal					
1:	8	11	1.00	1.00	
2:	6	9	0.92 (0.23-3.63)	0.00 (0.00-)	
3:	2	8	0.34 (0.06-2.07)	0.00 (0.00-)	
4:	4	8	0.69 (0.15-3.10)	6.18E+14 (0.00-)	1.00
At First Birth					
Pre-menopausal					
1:	18	53	1.00	1.00	
2:	15	43	1.03 (0.46-2.27)	2.69 (0.71-10.20)	
3:	14	44	0.94 (0.42-2.10)	1.00 (0.25-3.94)	
4:	17	33	1.52 (0.69-3.35)	1.59 (0.31-8.11)	0.22
Post-menopausal					
1:	14	42	1.00	1.00	
2:	31	30	3.10(1.41-6.80)*	11.31(2.49-51.50)*	

3:	22	27	2.44(1.07-5.59)*	3.45 (0.78-15.17)
4:	31	23	4.04(1.80-9.09)*	19.31(3.30-113.04)*	0.03*

* Odds ratios and 95% confidence intervals adjusted for age, education, race, BMI (1 yr before interview), age at first birth, age at menarche, age at menopause (for post-menopausal women only), number of births, first-degree relative with breast cancer, previous benign breast disease and year at interview.

Table 4-1. CYP1A1 and Risk of Breast Cancer

Genotype	Cases	Controls	Crude OR (CI)	Adjusted OR (CI)*
		At Menarche		
		Pre-menopausal		
1: wt	134	279	1.00	1.00
2: mutant	4	22	0.38 (0.13-1.12)	0.32(0.10-0.99)*
		Post-menopausal		
1: wt	42	60	1.00	1.00
2: mutant	2	5	0.57 (0.11-3.09)	0.98 (0.12-8.20)
		At First Birth		
		Pre-menopausal		
1: wt	133	294	1.00	1.00
2: mutant	4	24	0.37 (0.13-1.08)	0.33 (0.11-1.01)
		Post-menopausal		
1: wt	167	230	1.00	1.00
2: mutant	13	15	1.19 (0.55-2.58)	1.02 (0.43-2.44)

* Odds ratios and 95% confidence intervals adjusted for age, education, race, BMI (1 yr before interview), age at first birth, age at menarche, age at menopause (for post-menopausal women only), number of births, first-degree relative with breast cancer, and previous benign breast disease.

Table 4-2. CYP1A1 and Risk of Breast Cancer (whites only)

Genotype	Cases	Controls	Crude OR (CI)	Adjusted OR (CI)*
		At Menarche		
		Pre-menopausal		
1: wt	131	268	1.00	1.00
2: mutant	4	21	0.39 (0.13-1.16)	0.33 (0.11-1.02)
		Post-menopausal		
1: wt	40	54	1.00	1.00
2: mutant	2	5	0.54 (0.10-2.93)	1.08 (0.12-9.62)
		At First Birth		
		Pre-menopausal		
1: wt	129	280	1.00	1.00
2: mutant	4	23	0.38 (0.13-1.11)	0.34 (0.11-1.04)
		Post-menopausal		
1: wt	156	214	1.00	1.00
2: mutant	13	15	1.19 (0.55-2.57)	1.01 (0.42-2.42)

* Odds ratios and 95% confidence intervals adjusted for age, education, race, BMI (1 yr before interview), age at first birth, age at menarche, age at menopause (for post-menopausal women only), number of births, first-degree relative with breast cancer, and previous benign breast disease.

Table 5. Traffic PAH Exposure in Different Time Periods and Risk of Breast Cancer (CYP1A1 wt)

Quartiles of PAH Exposure	Cases	Controls	Crude OR (CI)	Adjusted OR (CI)*	P for trend
At Menarche					
Pre-menopausal					
1:	19	72	1.00	1.00	
2:	38	64	2.25(1.18-4.29)*	2.39 (0.96-5.93)	
3:	37	72	1.95(1.02-3.70)*	2.20 (0.84-5.73)	
4:	40	71	2.14(1.13-4.04)*	2.33 (0.91-5.96)	0.07
Post-menopausal					
1:	11	13	1.00	1.00	
2:	16	16	1.18 (0.41-3.41)	0.90 (0.18-4.60)	
3:	6	15	0.47 (0.14-1.64)	0.54 (0.09-3.43)	
4:	9	16	0.67 (0.21-2.09)	0.53 (0.09-3.18)	0.55
At First Birth					
Pre-menopausal					
1:	33	76	1.00	1.00	
2:	42	77	1.26 (0.72-2.19)	2.15 (0.90-5.14)	
3:	28	75	0.86 (0.47-1.56)	0.91 (0.34-2.40)	
4:	30	66	1.05 (0.58-1.90)	1.34 (0.43-4.13)	0.72
Post-menopausal					
1:	27	63	1.00	1.00	
2:	55	60	2.14(1.20-3.82)*	2.91(1.17-7.26)*	
3:	40	51	1.83 (0.99-3.38)	1.71 (0.64-4.58)	
4:	45	56	1.88(1.03-3.41)*	3.24(1.23-8.52)*	0.24

*   Odds ratios and 95% confidence intervals adjusted for age, education, race, BMI (1 yr before interview), age at first birth, age at menarche, age at menopause (for post-menopausal women only), number of births, first-degree relative with breast cancer, previous benign breast disease and year at interview.

Table 6. Traffic PAH Exposure in Different Time Periods and Risk of Breast Cancer (CYP1A1 mutant)

Quartiles of PAH Exposure	Cases	Controls	Crude OR (CI)	Adjusted OR (CI)*	P for trend
At Menarche					
Pre-menopausal					
1:	1	3	1.00	1.00	—
2:	1	10	0.30 (0.01-6.38)	/	
3:	2	4	1.50 (0.09-25.39)	/	
4:	0	5	0.00 (0.00-)	/	—
Post-menopausal					
1:	1	2	1.00	1.00	—
2:	1	1	2.00 (0.05-78.25)	/	
3:	0	2	0.00 (0.00-)	/	
4:	2	5	0.50 (0.00-)	/	—
At First Birth					
Pre-menopausal					
1:	0	6	1.00	1.00	—
2:	2	2	/	/	
3:	1	8	/	/	
4:	1	8		/	—
Post-menopausal					
1:	3	3	1.00	1.00	—
2:	3	3	1.00 (0.10-9.61)	/	
3:	3	6	0.50 (0.06-4.15)	/	

4:	4	3	1.33 (0.15-11.93)	/

* Odds ratios and 95% confidence intervals adjusted for age, education, race, BMI (1 yr before interview), age at first birth, age at menarche, age at menopause (for post-menopausal women only), number of births, first-degree relative with breast cancer, previous benign breast disease and year at interview.

CONCLUSION

Overall, our research activities have found evidence to support hypotheses that early life exposure to environmental pollutants may be associated with breast cancer risk. Specifically, we found evidence of geographic clustering of residence; the premenopausal women were more clustered than controls. The evidence for clustering of residential locations at birth and menarche was stronger than evidence for clustering at either the time of women's first birth or current residence. We also observed a general tendency of clustering of lifetime residence. Our findings suggest that there may be identifiable etiological processes on exposure and breast cancer risk, and that early exposures may be of particular importance.

Furthermore, we examined early life exposure to high concentrations of total suspended particulates, a surrogate for polycyclic aromatic hydrocarbons, in relation to breast cancer risk. We observed that exposure to high concentrations of total suspended particulates at birth was associated with an increase in risk of breast cancer for postmenopausal women. Conversely, in premenopausal women, the results were inconsistent with our hypothesis Early life exposure to environmental tobacco smoke was suggestive of a slight increase in the risk of breast cancer; however, we can not exclude the possibility that exposure was unrelated to risk.

Using a geographic model developed and validated by both Long Island and local data, we found some evidence of high traffic PAH exposure at menarche and first birth associated with increased risk of breast cancer; and this increased risk was only seen in non-smokers.

Genetic susceptibility analyses are currently in process. Our preliminary results suggest that GSTM1 null genotype is not associated with breast cancer risk; and high PAH exposure was only associated with increased breast cancer risk among women with GSTM1 null genotype.

LIST OF PERSONNEL RECEIVING
PAYMENT FROM GRANT

List of personnel receiving pay from the research effort:

Andriankaja, Oelisoa
Bordonaro, Rosaria
Freudenheim, Jo
Hill, Helyn
Johnson, Linda
LaFalce, Julie
Mark, David
McCann, Susan
Muti, Paola
Napoli, Diane
Nie, Jing
Rizzo, Diana
Saltino, Deborah
Vena, John
Vito, Dominica
Willett, Nicholas

KEY RESEARCH ACCOMPLISHMENTS

- We have identified and completed data entry for relevant industrial sites and major roadways during time periods under investigation.
- We have identified additional sources of information regarding historical sources of the exposures of interest and their locations and amounts.
- We have verified and geocoded residential histories of study participants.
- We have completed the geocoding for study participants for their residence at the time of their birth, at menarche, when they had a first birth and 10 and 20 years before diagnosis (cases) or interview (controls), approximately 20,000 addresses in Erie and Niagara counties.
- We have conducted a validation study of the positional accuracy of geocoded residences. Results of this validation study were published (1).
- We completed a GIS-based spatial and temporal analysis for residences of breast cancer cases and controls at early life and found strong evidence of spatial clustering for cases during this time. A manuscript has been accepted for publication (2) and abstracts have been presented (7, 9, 10, 12, 14, 15).
- We have completed data analysis examining early life exposure to total suspended particulates and exposure to environmental tobacco smoke in relation to risk of breast cancer in adult life. Two abstracts from this work were presented, one at the annual meeting of the Society for Epidemiologic Research, the other at the Annual Meeting of the Association for Cancer Research (6, 8). Manuscripts for this have been submitted for publication (3, 4).
- We have completed data analysis examining early life proximity to industrial sites contracted by the US Atomic Energy Commission in relation to risk of breast cancer in adult life. An abstract was presented (16).
- DNA extraction for approximately 3000 samples is close to completion and assessment of genotypes for four genes is underway.
- Doctoral dissertation (PhD), "Environmental Exposures in Early Life and the Risk of Breast Cancer," was completed April 3, 2003. Dr. Matthew Bonner is now a postdoctoral fellow in Environmental Epidemiology at NCI.
- Doctoral dissertation (PhD), "Geographical epidemiology of breast cancer in western New York: Exploring spatio-temporal clustering in GIS," was completed December 10, 2002. Dr. Daikwon Han is working on a postdoctoral fellowship funded by the USAMRMC.
- We completed updating of lifetime residential history of breast cancer cases and controls in western New York; all Erie and Niagara county residential location were identified and geocoded, and these were merged into one database. We checked consistency of geocoded addresses in different time points for each individual, updated incomplete addresses using Polk searches, and validated the consistency of reported years of moved in and out of the residence.
- We completed additional validation of the traffic model and further calibrated the model parameters. This traffic model, used to reconstruct historical traffic PAH exposure, was originally developed from the Long Island Breast Cancer Project.

- We completed data analysis examining traffic PAH exposure and risk of breast cancer in menarche, year of first birth, and year of exposure. A manuscript is being prepared.
- We completed preliminary data analysis examining PAH metabolic polymorphisms, GSTM1 and CYP1A1, in relation to breast cancer risk and gene environmental interaction. A manuscript is being prepared.
- A manuscript is in preparation regarding the clustering of lifetime residence and breast cancer risk using exploratory spatial analysis tools based on these lifetime residential history data.
- An abstract will be presented at the annual meeting of the International Society for Environmental Epidemiology in New York City, New York, August 2004.
- A paper was a semi-finalist in the Nystrom dissertation competition of the Association of American Geographers (AAG), and the paper was presented at the centennial meeting of the AAG, Philadelphia, PA. March 2004 (14).
- Doctoral dissertation (PhD), "Environmental Exposure to Polycyclic Aromatic Hydrocarbons (PAHs) Genetic Susceptibility and Risk of Breast Cancer. To be completed by September 1, 2004.

REPORTABLE OUTCOMES 2004

BIBLIOGRAPHY

Publications: (See APPENDICES I-IV)

1. "Positional Accuracy of Geocoded Addresses in Epidemiologic Research," Bonner, M., *Epidemiology*, 14-4, July 2003, 408-412.

2. "Geographic Clustering of Residence in Early Life and Subsequent Risk of Breast Cancer" Han, D, Rogerson, P, Nie, J, Bonner, M, Vena, J, Vito, D, Muti, P, Trevisan, M, Edge, S, Freudenheim, JL. *Cancer Causes and Control*, 2004 (In Press).

3. "Total Suspended Particulate Exposure in Early Life and Breast Cancer" Bonner, MR, Han, D, Nie, J, Vena, JE, Rogerson, P, Muti, P, Trevisan, M, Vito, D, Freudenheim, JL. Submitted for publication. (NOT FOR PUBLIC INFORMATION)

4. "Environmental Tobacco Smoke Exposure in Early Life and the Risk of Breast Cancer" Bonner, MR, Nie, J, Han, D, Vito, D, Vena, JE, Rogerson, P, Muti, P, Trevisan, M, Freudenheim, JL. Submitted for publication. (NOT FOR PUBLIC INFORMATION)

Abstracts/Presentations (See APPENDIX V)

5. "Validation of TIGER (Topologically Integrated Geographic Encoding and Referencing System) to Geocode Addresses for Epidemiologic Research," Nie, J, Bonner, MR, Vito, D, Willett, NH, Freudenheim, JL. Poster presentation, Congress of Epidemiology Meeting,Toronto, Canada, June 2001.

6. "Household Smoke Exposure in Early Life and Breast Cancer in Western New York" Bonner MR, Nie J, Vena JE, Rogerson P, Trevisan M, Freudenheim JL. American Association for Cancer Research, Toronto, Canada, April 2003.

7. "Clustering of Lifetime Residence and Breast Cancer Risk in Western New York" Annual Meeting of the Society for Epidemiologic Research, Atlanta, GA, June 2003.

8. "Total Suspended Particulate Exposure in Early Life and Breast Cancer," Bonner, M. Society for Epidemiologic Research, Atlanta, GE, June 2003.

9. "Geographical Epidemiology of Breast Cancer in Western New York" Annual Meeting of the Association of American Geographers in Los Angeles, CA, March 2002.

10. "Environmental Exposures associated with Lifetime Residential History: A GIS-based clustering Analysis of breast cancer," Annual Meeting of the Society for Epidemiologic Research, Palm Desert, CA, June 2002.

11. "Residential Proximity at Birth to Industrial Sites and Subsequent Risk of Breast Cancer," Bonner MR, Han D, Nie J, Freudenheim JL, Vena JE. American College of Epidemiology Annual Meeting, Albuquerque, NM, September 2002.

12. "Exploratory Spatial Analyses of Lifetime Breast Cancer Risk and Residence History," Han, D. Annual Meeting of the Association of American Geographers, New Orleans, LA, March 2003.

13. "Residential Proximity to Chemical or Primary Metal Industry and the Risk of Breast Cancer in Western New York," Nie, J, Bonner MR, Han D, LaFalce J, Vena J, Freudenheim JL. Society for Epidemiologic Research, Atlanta, GA, June 2003.

14. "Geographic Clustering of Residence in Early Life and Subsequent Risk of Breast Cancer," Han, Daikwon. Nystrom dissertation competition of the Association of American Geographers, Philadelphia, PA, 2004.

15. "Assessing the Variability of Risk Surfaces using Residential History Data in a Case Control Study of Breast Cancer," Han D, Rogerson, P, Bonner MR, Nie J, Vena J, Freudenheim JL. International Society for Environmental Epidemiology, New York, NY, 2004.

16. "Residential Proximity to Chemical or Primary Metal Industry and the Risk of Breast Cancer in Western New York," Nie, J, Bonner MR, Han D, LaFalce J, Vena JE, Freudenheim JL. Presented at the Annual Meeting of the Society for Epidemiologic Research, Atlanta, GAS, June 2003.

Degrees Obtained Supported by this Award:

Doctoral dissertation (PhD), "Environmental Exposures in Early Life and the Risk of Breast Cancer, completed April 3, 2003. (Matthew Bonner).

Doctoral dissertation (PhD), "Geographical epidemiology of breast cancer in western New York: Exploring spatio-temporal clustering in GIS," completed December 10, 2002. (Daikwon Han).

Doctoral dissertation (PhD), "Environmental Exposure to Polycyclic Aromatic Hydrocarbons (PAHs) Genetic Susceptibility and Risk of Breast Cancer. To be completed by September 1, 2004. (Jing Nie).

Traineeship Award:

Post-doctoral traineeship award, Department of Defense Breast Cancer Research Program, "Integrating Geographic Information System into Breast Cancer Epidemiologic Research" approved December 2002. (Daikwon Han).

APPENDICES

APPENDIX I

Positional Accuracy of Geocoded Addresses in Epidemiologic Research

Matthew R. Bonner,* Daikwon Han,† Jing Nie,* Peter Rogerson,† John E. Vena,* and Jo L. Freudenheim*

Background: Geographic information systems (GIS) offer powerful techniques for epidemiologists. Geocoding is an important step in the use of GIS in epidemiologic research, and the validity of epidemiologic studies using this methodology depends, in part, on the positional accuracy of the geocoding process.

Methods: We conducted a study comparing the validity of positions geocoded with a commercially available program to positions determined by Global Positioning System (GPS) satellite receivers. Addresses ($N = 200$) were randomly selected from a recently completed case–control study in Western New York State. We geocoded addresses using ArcView 3.2 on the GDT Dynamap/2000 U.S. Street database. In addition, we measured the longitude and latitude of these addresses with a GPS receiver. The distance between the locations obtained by these two methods was calculated for all addresses.

Results: The distance between the geocoded point and the GPS point was within 100 m for the majority of subject addresses (79%), with only a small proportion (3%) having a distance greater than 800 m. The overall median distance between GPS points and geocoded points was 38 m (90% confidence interval [CI] = 34–46). Distances were not different for cases and controls. Urban addresses (median = 32 m; CI = 28–37) were slightly more accurate than nonurban addresses (median = 52 m; CI = 44–61).

Conclusions. This study indicates that the suitability of geocoding for epidemiologic research depends on the level of spatial resolution required to assess exposure. Although sources of error in positional accuracy for geocoded addresses exist, geocoding of addresses is, for the most part, very accurate.

Key Words: geographic information systems, geocoding, address matching, epidemiology

Geographic information systems (GIS) are increasingly used by epidemiologists to screen and test hypotheses about environmental exposures and disease.[1-4] GIS techniques lend themselves to assessing residential or occupational proximity to exposures and to assessing spatial variation in epidemiologic measures. An important first step is often to geocode the study participant addresses.[4] Geocoding, also referred to as address matching, is the process whereby the relative positions of addresses are linked to a reference theme, which is a database that contains both address information and locational information (ie, latitude and longitude). A reference theme for geocoding, therefore, can be considered an electronic version of a street map. Geocoding is an attractive method for epidemiologists because GIS software is relatively inexpensive, uses routinely collected address data, and is very efficient at locating addresses. Verification of each subject's address location by other methods would require considerable time and resources, especially for a large study. Furthermore, verifying address locations is not feasible when subjects' lifetime series of addresses are considered because the number of these addresses can become extremely large.

The validity of epidemiologic studies using GIS and geocoding methods depends on the proportion of addresses that can be geocoded as well as the positional accuracy of the geocoding process. Several studies have assessed the address matching rate of commercial geocoding companies, and found that matching rates are typically 60–80%.[1,5] No previous published studies have assessed the positional accuracy of geocoding in epidemiologic research. Positional inaccuracy of geocoded addresses may be an important source of exposure misclassification in environmental epidemiology.

We describe here a study comparing the location of addresses measured by global positioning system (GPS) receivers (devices that use satellite signals to estimate the latitude and longitude of any given location) to positions geocoded with a commercially available reference theme. We assessed three areas that may be important to determine the appropriateness of geocoding in epidemiologic research. First, we compared the positional accuracy of historical addresses. Many exposures that are relevant to disease outcomes are historical in nature and positional inaccuracy of geocoding historical addresses may be a source of error in estimating these historical exposures. Second, we investigated whether positional inaccuracy of geocoded addresses would result in differential exposure misclassification between cases and controls. Third, we compared differences in positional accuracy between urban and nonurban areas. This urban–rural differential is particularly important because the reference themes commonly available were designed for nonepidemiologic purposes and are generally thought to be more accurate and complete in urban areas than in nonurban areas.

METHODS

We obtained a random sample of 200 addresses from a recently completed case–control study in Erie and Niagara Counties in Western New York State. Lifetime residential histories were collected from 3,286 subjects for a total of 20,240 addresses. These included study participants' current addresses and all previous addresses dating back to 1918. For the remainder of this article, addresses before the current address of each participant are termed "historical addresses." Most addresses (N = 15,903) were for residences in Erie and Niagara Counties. We geocoded Erie and Niagara County addresses using ArcView 3.2 (ESRI, Inc., Redlands, CA) and the Dynamap/2000 US Street Database (Geographic Data Technologies, Inc., Lebanon, NH) for Erie and Niagara Counties as the reference theme. Essentially, the Dynamap/2000 is an enhancement of the Topologically Integrated Geographic Encoding and Referencing file (TIGER/line file) that was developed by the US Bureau of the Census. These are data files that contain street address ranges and census tract/block boundaries.[6] We matched 10,356 (65%) of the original 15,903 addresses using the initial geocoding process.

To ensure an adequate number of cases and controls for comparison of urban and nonurban positional accuracy, we randomly selected 200 addresses in a random block selection scheme to obtain 50 cases and 50 controls from urban areas and 50 cases and 50 controls from nonurban areas. We defined urban addresses as addresses within the city limits of Buffalo, Niagara Falls, and Kenmore, NY. All other addresses were considered nonurban. If an address could not be located for the GPS measurements, then that address was discarded and a new address was randomly selected from the same block.

We determined the geocoded latitude and longitude for each address with ArcView 3.2 by first geocoding the addresses and then changing the map projection to Universal Transverse Mercator-1983. This projection compensates for the Earth's curvature and generates more precise estimates of latitude and longitude than other projections. Latitude and longitude were then converted into x and y coordinates (arbitrary values representing a point on a plane) for each address. These x and y coordinates were measured in meters for this study.

We measured the actual geographic position of the 200 addresses with an Etrex GPS receiver manufactured by Garmin (Garmin International, Inc., Olathe, KS). This GPS receiver reported latitude and longitude in decimal degrees to five places using the World Geodetic System 1984 map datum. Before making site visits, the GPS receiver was turned on and automatically searched for least three satellite signals. Once the satellite signals were detected, the GPS receiver provides real-time current latitude, longitude, speed, and direction. Investigators then visited each address and used the GPS receiver to record latitude and longitude from the street directly in front of each address.

The observed GPS latitudes and longitudes were then converted to x and y coordinates in Universal Transverse Mercator-1983 projection for comparison between the GPS and the geocoded positions. We calculated the distance between the two points for each address by determining the Euclidean length of the hypotenuse of the right triangle formed by the two points: $[(x_1 - x_2)^2 + (y_1 - y_2)^2]^{1/2}$, where x_1 is the GPS latitude, x_2 is the geocoded latitude, y_1 is the GPS longitude, and y_2 is the geocoded longitude. This formula does not correct for the curvature of the Earth. However, this uncorrected formula does not introduce sizable error in the distance calculation between the two points because the distances between the points were relatively small.

The mean distance, its standard deviation, and the median distance between the geocoded points and the GPS points were calculated with SPSS version 10.1 for the total sample, for case and control addresses, for urban and nonurban addresses, and for cases and controls stratified by urban/nonurban status. We grouped distance into nine categories: <25 m, 25–50 m, 51–75 m, 76–99 m, 100–199 m, 200–399 m, 400–599 m, 600–799 m, and >800 m; the proportion of addresses in each of these categories was then computed. Bootstrapped 90% confidence intervals for the medians of distance were computed with Resampling Stats (Resampling Stats, Inc., Arlington, VA).

RESULTS

The majority of the 200 randomly selected addresses (N = 133) were historical in nature; subjects did not currently occupy these addresses (Table 1). The median distances

TABLE 1. Distance (m) Between Geocoded Position and GPS Position for 200 Addresses, by the Year Moved Out of Address

Year Moved Out of Address*	% of Addresses (N = 200)	Median Distance (90% CI)	Minimum Distance	Maximum Distance
1930–1939	1.5	33 (16–50)	16	50
1940–1949	8.0	40 (32–53)	17	1225
1950–1959	11.5	38 (26–53)	9	502
1960–1969	11.5	29 (26–38)	9	760
1970–1979	14.5	36 (29–71)	7	2552
1980–1989	12.0	38 (28–59)	5	313
1990–2000	7.5	34 (28–70)	12	209
Currently occupy	32.5	49 (38–59)	6	763
Unknown	1.0	96 (19–172)	19	172
Total	100	38 (34–46)	5	2552

GPS = global positioning system.
*Indicates the decade when a study participant moved out of the address.

between the GPS and the geocoded position did not vary greatly across decades. For the current addresses, there was a slightly larger median distance (49 m; 90% confidence interval [CI] = 38–59) between the measured address location and the geocoded location, whereas the distances for the historical addresses tended to be between 30 and 40 m. However, the three addresses with distances greater than 1,000 m were all historical addresses.

Tables 2 and 3 present comparisons of the geocoded

TABLE 2. Distance Between Geocoded Position and GPS Position for Case and Control Addresses

	Cases (N = 100)	Controls (N = 100)	All Addresses (N = 200)
Distance (m)			
Median	41	38	38
90% CI	31–47	33–49	34–46
Mean (SD)	119 (247)	107 (286)	113 (266)
Minimum distance	5	5	5
Maximum distance	1151	2552	2552
Distance (%)			
<25 m	28	26	27
25–50 m	33	34	33.5
51–75 m	12	16	14
76–99 m	5	4	4.5
100–199 m	10	9	9.5
200–399 m	5	8	6.5
400–599 m	1	1	1
600–799 m	2	0	1
≥800 m	4	2	3

GPS = global positioning system; SD = standard deviation.

and GPS positions. The distribution of distances between geocoded and GPS positions is skewed to the right, as evidenced by the median distance for all addresses (38 m; 90% CI = 34–46) being considerably smaller than the mean distance (113 m; Table 2). Consequently, the median is more accurate than the mean as a measure of central tendency. The distance between the geocoded point and the GPS point was within 100 m for the majority of the all subject addresses (79%), with only a small proportion (3%) having a distance greater than 800 m. Distances were not different for cases and controls.

Positional accuracy was somewhat better for the urban addresses (32 m; 90% CI = 28–37) than for the nonurban addresses (52 m; 90% CI = 44–61) (Table 3). In addition to having a smaller median distance, urban addresses had a higher proportion of addresses within 100 m (89%) than the nonurban addresses (69%). Within the urban strata, cases and controls were more similar than the cases and controls in the nonurban strata. In the nonurban strata, there was a 9-m difference in the medians between cases (45 m; 90% CI = 40–70) and controls (54 m; 90% CI = 38–66). In addition, urban cases (86%) and controls (92%) had a higher proportion of addresses within 100 m than did nonurban cases (70%) and controls (68%).

DISCUSSION

Overall, the positional accuracy of addresses geocoded with the Dynamap/2000 was good. The majority of addresses were located within 100 m of the real address as determined by on-site GPS latitude and longitude measurements; the historical addresses tended to have smaller median distances than the current addresses. There was, however, some difficulty in assessing all the selected historical addresses. In eight

TABLE 3. Distance Between Geocoded Position and GPS Position for Case and Control Addresses, by Urban/Nonurban Residential Status

	Urban			Non-Urban		
	Cases (N = 50)	Controls (N = 50)	Total (N = 100)	Cases (N = 50)	Controls (N = 50)	Total (N = 100)
Distance (m)						
Median	31	32	32	45	54	52
90% CI	27–41	28–37	28–37	40–70	38–66	44–61
Mean (SD)	122 (285)	70 (177)	96 (237)	116 (204)	125 (362)	129 (293)
Minimum distance	6	5	5	5	12	5
Maximum distance	1551	1225	1551	1223	2552	2552
Distance (%)						
<25 m	36	30	33	20	22	21
26–50 m	34	46	40	32	22	27
51–75 m	14	12	13	10	20	15
76–99 m	2	4	3	8	4	6
100–199 m	4	2	3	16	16	16
200–399 m	0	4	2	10	12	11
400–599 m	2	0	1	0	20	10
600–799 m	2	0	1	2	00	10
≥800 m	6	2	4	20	20	20

GPS = global positioning system.

instances, subjects' homes appeared to have been demolished, leaving a vacant lot. The uncertainty about whether these lots were the correct street address prevented GPS measurements of these addresses. This indicates that in areas where there has been major redevelopment or neglect the positional accuracy of historical addresses may be more difficult to determine.

Positional accuracy was not different between cases and controls, suggesting that errors resulting from the geocoding of addresses may not result in differential misclassification of exposure. In addition, the difference in positional accuracy between urban addresses and nonurban addresses was small. However, even with good overall positional accuracy, there were several sources of error. First, the Dynamap/2000 is largely derived from the US Bureau of the Census TIGER/line files, and inaccurate methods were used to create these TIGER/line files. The Geography Division of the US Bureau of the Census has reported that the median distances between GPS measured positions and the TIGER/line file positions in eight US counties ranged between 30 and 121 m.[7] Additionally, this report also indicated that TIGER/line file updates since 1990 are less accurate than the pre-1990 versions. These inaccuracies may have important implications for epidemiologic use because, as the updates to the TIGER/line files provide more complete coverage and increase the address-matching rate, the decrease in positional accuracy may lead to increased error and misclassification of

exposure when estimating exposures based on geographic positioning.

The second source of error with geocoding arises from the geocoding process itself. We found that positional accuracy was slightly decreased in nonurban addresses compared with urban addresses, likely a result of the geocoding process. Geocoding uses interpolation to estimate the relative position of an address on a line segment in the reference theme.[4,8] The likelihood of inaccurate interpolation by the geocoding process is higher for an address location in areas with long street segments than in areas where there are many short street segments, regardless of urban/nonurban location; however, nonurban areas generally have longer street segments than urban areas.

In addition to error in positional accuracy from the Dynamap/2000 and geocoding, GPS receivers are also prone to error. These errors are generally small but remain a limitation of this study in that the GPS receiver was used as the standard. GPS errors in positional accuracy arise from three general sources: satellite related errors, signal propagation errors, and receiver errors. Garmin reports that the Etrex has positional accuracy between 1 and 5m. Field tests, however, indicate that civilian GPS receivers are only accurate within 15–40 m.[9] The Dynamap/2000 may actually be more accurate than we report here because most of the distances between the GPS point and the geocoded point were within the range of error for the GPS unit.

These errors in the positional accuracy of geocoding study participant addresses and sources of exposure have several implications for epidemiologic research. First, numerous epidemiologic studies have crudely defined exposure based solely on proximity of a residential address to an exposure of interest. For instance, McLaughlin and associates[10] used a 25-km radius around nuclear facilities in Canada to classify those residing within that circle as exposed and those outside as unexposed. Clearly, geocoding has sufficient spatial resolution to distinguish differences on this scale. In another study, Croen et al[11] investigated maternal residential proximity to hazardous waste sites and congenital malformations. Maternal residence within 1 mile (1.6 km) of a hazardous waste site was defined as exposed. Again, even with the higher required spatial resolution to define exposure, geocoding should have sufficient positional accuracy to classify exposure appropriately.

There are exposures, however, with great spatial variation over relatively small distances. When these exposures are being considered, the study may require more precise methods to locate subjects and exposure sources to produce valid risk estimates. Electromagnetic fields, for instance, require high spatial resolution to estimate exposure accurately. The intensity of magnetic fields decreases exponentially with distance. In many epidemiologic studies that investigate electromagnetic fields and cancer, residential proximity to power lines and electricity transmission equipment has been measured in meters rather than kilometers or miles as in the previous examples. For example, in their meta-analysis of residential proximity to electricity transmission and distribution equipment and childhood cancer, Washburn et al[12] used less than 50 m to define the exposed group. Based on the present validation of geocoding, there is some question whether the positional accuracy of geocoding is sufficient. Use of geocoding in situations where high spatial resolution is required may lead to extensive nondifferential misclassification of exposure, thereby greatly reducing the validity of the risk estimates.

The generalizability of this study may be limited by regional variation in the completeness of the Dynamap/2000. These results should be comparable to regions where the Dynamap/2000 has completeness similar to that of Erie and Niagara Counties. Furthermore, as the US Bureau of the Census and GDT continue to develop and improve the TIGER/line and the Dynamap/2000 databases, regional variation will diminish. However, recent improvements in the completeness of these updated street databases may not alleviate the concerns about their positional accuracy because post-1990 updates have not been as carefully assembled as the pre-1990 updates.

Finally, as GIS becomes more commonly used in epidemiologic research, the need to assess geocoding methods and reference themes will become more important, especially when high spatial resolution is required to classify a study subject's exposure accurately. Overall, this study indicates that these tools are sufficiently accurate for some -but not all- epidemiologic studies. Consequently, care should be taken in the interpretation of results, taking into account sources of error in positional accuracy for geocoded addresses that may affect exposure classification.

ACKNOWLEDGMENTS

REFERENCES

1. Krieger N, Waterman P, Lemieux K, et al. On the wrong side of the tracts? Evaluating the accuracy of geocoding in public health research. *Am J Public Health*. 2001;91:1114–1116.
2. Moore DA, Carpenter TE. Spatial analytical methods and geographic information systems: use in health research and epidemiology. *Epidemiol Rev*. 1999;21:143–161.
3. Bellander T, Berglind N, Gustavsson P, et al. Using geographic information systems to assess individual historical exposure to air pollution from traffic and house heating in Stockholm. *Environ Health Perspect*. 2001;109:633–639.
4. Vine MF, Degnan D, Hanchette C. Geographic information systems: their use in environmental epidemiologic research. *Environ Health Perspect*. 1997;105:598–605.
5. Howe HL. Geocoding NY State Cancer Registry. *Am J Public Health*. 1986;76:1459–1460.
6. Anonymous. *Getting to Know ArcView GIS*. 3rd ed. Redlands, CA: Environmental Systems Research Institute, Inc.; 1999.
7. Liadis JS. *GPS TIGER Accuracy Analysis Tools (GTAAT) Evaluation and Test Results*. United States Bureau of the Census, Division of Geography, TIGER Operations Branch, 2000.
8. Drummond WJ. Address matching:GIS technology for mapping human activity patterns. *J Am Planning Assoc*. 1995;61:240–251.
9. Hofmann-Wellenhof B, Lichtenegger H, Collins J. *Global Positioning Systems*: Theory and Practice. New York: Springer-Verlag; 1997.
10. McLaughlin JR, Clarke EA, Nishri ED, Anderson TW. Childhood leukemia in the vicinity of Canadian nuclear facilities. *Cancer Causes Control*. 1993;4:51–58.
11. Croen LA, Shaw GM, Sanbonmatsu L, et al. Maternal residential proximity to hazardous waste sites and risk for selected congenital malformations. *Epidemiology*. 1997;8:347–354.
12. Washburn EP, Orza MJ, Berlin JA, et al. Residential proximity to electricity transmission and distribution equipment and risk of childhood leukemia, childhood lymphoma, and childhood nervous system tumors: systematic review, evaluation, and meta-analysis. *Cancer Causes Control*. 1994;5:299–309.

APPENDIX II

Geographic Clustering of Residence in Early Life and Subsequent Risk of Breast Cancer

Daikwon Han[1,2], Peter A. Rogerson[2], Jing Nie[1], Matthew R. Bonner[1], John E. Vena[3], Dominica Vito[1], Paola Muti[1], Maurizio Trevisan[1], Stephen B. Edge[4], Jo L. Freudenheim[1]

[1] Department of Social and Preventive Medicine, University at Buffalo, Buffalo, NY 14214.

[2] Department of Geography and National Center for Geographic Information and Analysis, University at Buffalo, Buffalo, NY 14261.

[3] Department of Epidemiology and Biostatistics, University of South Carolina, Columbia, SC 29208.

[4] Department of Breast and Soft Tissue Surgery, Roswell Park Cancer Institute, Buffalo, NY 14263.

Corresponding author: Daikwon Han, Department of Social and Preventive Medicine, 270 Farber Hall, University at Buffalo, Buffalo, NY 14214
Phone: 716-829-2975 x 605, Fax: 716-829-2979, Email: dhan@buffalo.edu

Abstract

Objective: This study focused on geographic clustering of breast cancer based on residence in early life and identified spatio-temporal clustering of cases and controls.

Methods: Data were drawn from the WEB study (Western New York Exposures and Breast Cancer Study), a population-based case-control study of incident, pathologically confirmed breast cancer (1996-2001) in Erie and Niagara counties. Controls were frequency-matched to cases on age, race, and county of residence. All cases and controls used in the study provided lifetime residential histories. The k-function difference between cases and controls was used to identify spatial clustering patterns of residence in early life.

Results: We found that the evidence for clustered residences at birth and at menarche was stronger than that for first birth or other time periods in adult life. Residences for premenopausal cases were more clustered than for controls at the time of birth and menarche. We also identified the size and geographic location of birth and menarche clusters in the study area, and found increased breast cancer risk for pre-menopausal women whose residence was within the cluster compared to those living elsewhere at the time of birth.

Conclusion: This study provides evidence that early environmental exposures may be related to breast cancer risk, especially for pre-menopausal women.

Key words: spatial clustering, breast cancer, early-life exposure

Introduction

Breast cancer is one of the leading causes of death among women in the United States. However, the epidemiology of breast cancer is not yet fully understood. We also do not fully understand mechanisms for the known risk factors; for instance, why changes in age at menarche or age at first birth have an impact on breast cancer risk. A substantial degree of geographical variation in breast cancer incidence and mortality in the US has been observed.[1,2] While inconclusive, several environmental risk factors are also believed to be involved in breast cancer incidence.[3,4] There is speculation that environmental factors may explain geographic variation in breast cancer rates not explained by known risk factors. For this reason, the potential role of environmental exposures in breast cancer risk is of particular interest.

In addition, there is a growing interest in early life and lifetime exposures in relation to breast cancer risk. The life course approach is of interest because there may be sensitive time periods for exposures and/or there may be cumulative effects of lifetime exposure involved in breast cancer incidence.[5,6] Early life has an effect on breast cancer etiology evidenced by the known risk factors such as age at menarche, age at first birth and parity. There is new evidence that even earlier exposures may have an impact on adult breast cancer risk.[7] Trichopoulos[8] suggested that the *in-utero* and perinatal period might be pathologically significant and that the risk of adult breast cancer could be related to high estrogen exposure in early life. There is also accumulating evidence that factors related to early exposure, such as birthweight, may be related to risk.[9,10]

There has been little research investigating possible effects of environmental exposures in early life on subsequent breast cancer risk. Using residence as a proxy

11

measure for environmental exposures, we investigated whether there was any evidence of geographic clustering of adult breast cancer cases associated with their residences in early life. Clustering analyses have often been used to provide clues for the unknown etiology of disease, and thus to generate hypotheses for further epidemiologic research.[11] We looked at the geographic clustering of residence at early critical time points; at birth, at menarche, and at the woman's first birth. By comparing differences in clustering patterns between case and control residences, we were interested in identifying time periods critical to potential environmental exposures and subsequent breast cancer risk.

Methods

Population-based case-control study of breast cancer

We conducted a case-control study of breast cancer in western New York -- the WEB study (Western New York Exposures and Breast Cancer Study). Cases were women, age 35-79 with incident, primary, pathologically confirmed breast cancer diagnosed in Erie and Niagara counties during the period 1996-2001, with no previous cancer diagnosis other than non-melanoma skin cancer. Controls were frequency matched to cases on age, race, and county of current residence; controls under 65 years of age were randomly selected from a New York State Department of Motor Vehicles list and those 65 years and over were chosen from a Health Care Finance Administration list. We ascertained cases by having a nurse-case finder visit the pathology departments of almost all hospitals in these counties. One hospital which did not participate does almost no cancer surgery and refers patients to other participating hospitals. For the one other hospital which did not participate, breast cancer cases were identified in the practice of the breast surgeons who

see more than 99% of the cases from that hospital. Extensive in-person interviews and self-administered questionnaires were used to ascertain lifetime residential history and other breast cancer risk factors. A total of 1166 cases and 2105 controls were interviewed. Response rates were 72% and 65% for cases and controls, respectively.

All participants were asked to complete a lifetime residential history, to list the street address, town/city and zip code for their current address and then all other previous addresses throughout their lifetime. Participants provided 20,240 addresses, an average of approximately six addresses for each individual. In this study we focused on residence at the time of the participants' birth, menarche, and at the time that she had her first birth. Analyses were restricted to women residing in Erie or Niagara counties at each of these time points. There were, of course, participants whose addresses were the same for two or more of these times.

For women with incomplete residential information, additional information was obtained using historical city directories. We used these directories to find old addresses, and utilized various resources, such as web searches and commercial address databases for recent addresses. We also examined validity and reliability of reports of earlier residences in a number of ways. For birth addresses, we asked for information on birth address twice and have collected information on reliability of response. For the other time periods, we used information on maiden name and partial address information provided by the participants to search for records in city directories for the appropriate time periods. To improve our ability to geocode addresses, we developed several strategies. First, all addresses were standardized to be matched with the standard format used in GIS. We used the enhanced version of TIGER (Topologically Integrated Geographic Encoding

13

and Referencing Systems), GDT/Dynamap 2000[12], and overall matching rates were improved about 15-20 % when compared with the use of TIGER as a reference theme. We also used the stand-alone address cleaner ZP4 (Semaphore Co.) to correct and update zip code information to be matched with United States Postal Services certified addresses.

More than 85% of addresses were geocoded using the above strategies and resources. We failed to geocode some addresses primarily because of missing residential information, such as missing street numbers or street names. Since we are dealing with historical residential information, the likelihood of missing previous residential information was higher than that for current residential information. Table 1 is a summary table showing the numbers of cases and controls with complete residential information who resided in the two counties for each of the time periods. The percentage of missing residential information associated with each early life event was highest for birth addresses, at about 20%.

Clustering analyses of residences

To compare clustering patterns of breast cancer cases and controls at each time period, the primary method used was based on the k-function.[13] The k-function for a point process is defined as the number of events within distance h of an arbitrary event, divided by the overall intensity of events. It is estimated by

$$\lambda \hat{k}(h) = \sum_{i=1}^{n} \sum_{j=1}^{n} w(s_i, s_j)^{-1} I(d_{ij} \leq h)/n, \ \ h > 0$$

where n is the number of events, λ is the expected density of events in the study region, h is the pre-specified distance, d_{ij} is the Euclidian distance between point i and point j, I is an indicator function that is equal to one if inter-event distances (d_{ij}) are less than or equal

14

to h, and zero otherwise, and $w(s_i, s_j)$ is an edge correction estimator which is the proportion of the circumference of a circle centered at s_i, passing through s_j and that is inside the study area A.[14] Under the null hypothesis of spatial randomness, the expected value of $k(h)$ is πh^2 Geographic clustering will yield values of the k-function that are greater than this, since clustering will result in more pairs of points separated by a distance of h than would be expected in a random pattern.

We used the difference between k-functions for cases and controls to compare two patterns (i.e., $D(h) = k_{case}(h) - k_{control}(h)$). Positive values of $D(h)$ indicate spatial clustering of cases relative to the spatial clustering of controls. Under the null hypothesis of random labeling of cases and controls, the expected value of $D(h)$ is zero, indicating that the k-functions of the cases and controls are the same. The test statistic, $D(h)$, was calculated with confidence envelopes using the *splancs* library in *S-plus*.[15] We obtained the approximate 95% confidence limits for two standard errors ($\pm 2\sqrt{[Var\{D(h)\}]}$) at the $\alpha = .05$ level.[16] When the estimated function $D(h)$ deviated from zero by greater than two standard deviations, we interpreted this as a statistically significant difference between the case and control patterns.

We also employed a spatial clustering method to identify significant geographic clusters of breast cancer cases. The spatial scan statistic[17], which considers the likelihood of observing the actual number of cases inside of the circle under the null hypothesis of no clustering, was applied to residence at early life events. We were mainly interested in spatial clustering of high rates, and employed the Bernoulli model based on the locations of individual cases and controls.[18] In addition, odds ratios (OR) and 95% confidence intervals (95% CI) were obtained using logistic regression, adjusting for age, education,

age at menarche, parity, history of benign breast disease, family history of breast cancer. All analyses were conducted for the entire group of study participants and for data stratified on menopausal status. Women were considered postmenopausal if their menses had ceased permanently and naturally. Among other women, participants were also considered postmenopausal if any of the following conditions were true; they were on hormone replacement therapy and were over age 55, they had had a bilateral oophorectomy, they had had a hysterectomy without removal of the ovaries and they were older than 50, their menses had ceased permanently due to radiation or other medical treatment and they were older than 55.

Results

Characteristics of subjects included in the analysis, subjects with missing residential information, and subjects excluded due to residence outside of Erie and Niagara counties, are shown in Table 2. About half of the sample was excluded for each time period; the highest percentage of ineligible cases and controls was at the birth residence (46% and 51% respectively). However, we found little difference in characteristics between those subjects included and those subjects with addresses outside of these two counties.

Mapping was used to identify geographic patterns of breast cancer cases and controls for each of the early life events. Maps showing the locations of cases and controls in Figure 1 portray the underlying geographic patterns of breast cancer cases and controls in the study area. The rectangular region was used instead of the actual county boundary as an approximate boundary of the study area to protect individuals'

confidentiality. The purpose of such mapping is to inspect patterns visually -- the first step in any spatial analysis. Geographic patterns do not appear to vary much from one time period to the next, and they appear to reflect patterns of population distribution in the study area. However, it is difficult to determine whether they were clustered or dispersed relative to population from visual inspection alone, because of the large number of data points.

To assess potential effects of geographic selection bias in our study, we also examined the distribution of current residence in relation to other population data on geographic distribution of breast cancer cases and the general population. We did not find differences in the geographic distribution of participating and non-participating cases, nor between controls and the underlying population, except some tendency for both cases and controls living closer to the interview site to be somewhat more likely to participate than those living further away.

Spatial clustering of residences associated with early life events

We obtained differences between the case and control patterns for locations associated with each early life event. The k-function differences for values of h up to 15 miles, with approximate 95 % confidence envelopes, are shown in Figure 2. The maximum value of h is generally taken as one-third of the linear extent of the study area.[19] Any patterns beyond this scale can be disregarded, since either peaks or troughs in this geographic scale are difficult to interpret, and are potentially misleading. Figure 2a shows k-function differences for birth residence. It is clear that the estimated function shows strong evidence of spatial clustering, that is, of clustering of cases relative to

17

controls. There was no significant difference up to 3 miles; statistically significant differences were detected beyond the scale of 3 miles. There is also evidence of some degree of clustering for breast cancer cases at menarche residence (Figure 2b). Estimates of the D-function are positive but not statistically significant up to 7 miles; spatial clustering of breast cancer cases occurs at a scale of about 7-15 miles. For residence at women's first birth and for current residence, the difference is not statistically significant; the plot falls within the confidence interval over all distances (Figures 2c and 2d).

To determine whether there are any differences in clustering patterns by menopausal status, the k-function difference was performed for premenopausal and postmenopausal women separately (Figure 3). We found significant clustering of premenopausal breast cancer cases compared to controls for both birth and menarche residence (Figures 3a), while there is no evidence of clustering for postmenopausal breast cancer cases for either period (Figures 3b). We did not find evidence of clustering for first birth and current residence (at diagnosis) for either group (not shown). Estimated functions at birth residence show a strong clustering of premenopausal cases over the entire geographic scale with a peak at 7 miles. Values are positive for post-menopausal cases, but not statistically significant. For menarche residence, we also observed a strong clustering of premenopausal cases with a peak at about 8-10 miles. Again differences are not statistically significant for postmenopausal women at menarche residence.

Identifying the geographic location of breast cancer clusters

To identify the geographic location of areas with higher intensities for pre-menopausal cases in the study area, the spatial scan statistic was applied to residences of

premenopausal women at the time of birth and menarche. Maps in Figure 4 present results of the clustering analysis. The circle in Figure 4a indicates clustering of birth residence for premenopausal cases when compared to controls. We found a circular cluster of birth residence for breast cancer cases with a 5.7 mile radius in the area including part of the city of Buffalo, and the towns of Amherst, Cheektowaga, and Tonawanda (shaded areas). There are 100 observed breast cancer cases inside the cluster, while 76 breast cancer cases are expected. The cluster was significant at <0.01 with 999 Monte Carlo simulations.

Further, we examined breast cancer risk associated with residence in the cluster at the time of birth. When we compared other breast cancer risk factors, such as age, education, and age at menarche, for the premenopausal breast cancer cases and controls whose birth residence was inside the cluster to those who lived outside of cluster, we did not find significant differences between the two groups (data not shown). We observed an elevated breast cancer risk for premenopausal women living in the cluster at the time of birth. With subjects living outside the cluster as a reference group, the adjusted odds ratio was 2.65 (95% CI 1.75-4.0) after controlling for age, education, age at menarche, parity, history of benign breast disease, and family history of breast cancer.

We also identified clustering of menarche residence for premenopausal women and obtained similar results as for birth residence. We were able to identify a small clustering of menarche residences for premenopausal breast cancer cases. A small cluster in the center of those four towns was detected (Figure 4b). It is a small-sized cluster with 0.8 mile radius and is statistically significant at $p<0.05$. The cluster contains 9 observed and 3.1 expected breast cancer cases, yielding a relative risk (ratio of observed to

expected breast cancer cases) of 2.9. A secondary cluster was also detected near the city of Buffalo. It has a three mile radius and relative risk of 1.38 with 65 observed and 47 expected breast cancer cases, but it is not statistically significant (p=0.38).

Discussion

To our knowledge, no other studies have examined clustering of residential locations associated with cancer during early life: studies have examined clustering of residential locations at the time of diagnosis or death.[20] Critical time periods, including birth, menarche, and women's first pregnancy, as important early life and reproductive events in women's life, may play a substantial role in the risk of breast cancer. Under the hypothesis that there may be sensitive time periods in women's lives that will carry greater risk for exposure, the essential question was whether cases were more clustered than the underlying population, as represented by the controls. We found that cases were more clustered than controls at the time of birth and menarche, and it was due to clustering of residence for pre-menopausal, but not for post-menopausal breast cancer. The evidence for clustering of residential locations at birth and menarche was stronger than evidence for clustering at the time of women's first birth or other time periods in adult life. Our findings suggest that there may be identifiable etiological processes linking exposure and breast cancer risk, especially for premenopausal women, and that early exposures may be of particular importance.

This study provided a unique opportunity to examine clustering of breast cancer cases and controls at various points during early life. The facts that the study area had a

relatively stable population and about forty percent of study participants were lifetime residents, made the results more reliable. The evidence that residence in early life was important in the geographical clustering of breast cancer cases may be of particular importance for understanding environmental determinants of breast cancer. These findings suggest the importance of early or lifetime exposure in relation to disease risk in adult life, and also the potential role of the effects of migration on exposures and disease risk. Although migration can have a serious effect on the detection of geographical differences in disease risk, it has not been adequately addressed in previous clustering analyses.[21] Further investigations are required to prove any relationship between geographic clustering of residence and breast cancer risk, and the effects of residential changes on exposures should be considered in these studies.

Our finding of clustering was restricted to premenopausal breast cancer. We stratified on menopausal status because of evidence that there were differences in risk factors for pre- and postmenopausal women.[22] The mechanism of the observed difference is not clear. It could be that early life exposures impact premenopausal more than postmenopausal disease because of greater temporal proximity. There is some evidence, though not consistent, that other early exposures may differ by menopausal status. For example, there are data suggesting that birthweight may be more associated with pre- than with postmenopausal breast cancer.[9, 23]

The results should be interpreted cautiously due to the fact that there may be some artifacts of the analysis. First, it is important to note that spatial point patterns are complex to summarize in a single way.[24] For example, the use of cumulative scales in the application of the k-function method may influence the outcome.[25] In particular,

clustering is more likely to be detected on a larger geographic scale, and it tends to show continuous patterns over several neighboring scales due to the fact that the geographical scales are cumulative. Further refinement of methods to summarize spatial point patterns may provide more reliable results, as well as more accurate estimates of disease risk.

Second, this study is limited to current residents in the study area because we focused on the residential environment of Erie and Niagara counties; participants residing outside of these two counties at the time of each early life event were not included. The existence of missing residential information and potential selection bias due to non-participation may influence the results. As noted, we found no difference in participation by residence for cases compared to controls. Further we would expect that our findings on the clustering of early-life residence would be less subject to potential geographic selection bias than would current residence. We found a greater degree of clustering for residence at early life, than for current residential location.

Further, the fact that residence at birth and menarche were often the same made it difficult to differentiate associations for the two time periods. For 22% of cases and 35% of controls, the menarche residence was the same as their birth residence. While the observed tendencies may be related to environmental exposures, it is also possible that clustering of residence at the time of birth or menarche may be due to clustering of other socioeconomic or demographic factors. Evaluation of the contribution of socioeconomic status to clustering of residences at birth and menarche is of special interest. There may be other factors associated with residence not measured in this study. The findings are still of interest for further study in order to understand what those exposures might be. We are now investigating the relation between spatio-temporal clustering of residences

22

and exposures to environmental compounds, such as PAHs and benzene, to provide epidemiologic evidence of this finding.

Since the publication of John Snow's[26] well-known cholera map for the city of London in the 19th century, the relationship between the environment and disease has been one of the major research themes in medical geography. Geographic perspectives are of great use in describing geographical patterns of diseases, generating hypotheses on disease etiology, monitoring high risk areas of disease incidence, and suggesting possible causal factors of particular disease.[27, 28] Our study demonstrated that these GIS-based clustering analyses provide effective ways to explore spatial-temporal patterns of clustering. The findings show consistent results; the cluster identified by spatial analyses remained significant when traditional epidemiologic methods were used, and it was not explained by potential confounders. A recent study comparing "traditional" epidemiological methods, GIS, and point pattern analysis for use in the spatially referenced public health data concluded that results complement, rather than contradict or duplicate each other.[29]

In summary, this analysis of breast cancer clustering in space provides evidence of geographic clustering of premenopausal, but not postmenopausal, breast cancer cases at the time of birth and menarche, suggesting a possible influence of exogenous risk factors on breast cancer at these time points. While it is not clear from these data what caused this spatial clustering, it is provocative in providing evidence of the importance of this early period in breast carcinogenesis. Further investigations on genetic susceptibility may be of relevance to identify different effects on pre- and postmenopausal breast cancer. It will also be meaningful to see whether there is temporal clustering of early-life

23

residences as well as spatial clustering. This type of study also needs to be replicated in other settings.

References

1. Sturgeon SR, Schairer CG, McAdams M, Brinton LA, Hoover RN (1995) Geographic variation in mortality from breast cancer among white women in the United States. J Natl Cancer Inst 87: 1846-1853.

2. Lacey JV Jr, Devesa SS, Brinton LA (2002) Recent trends in breast cancer incidence and mortality. Environ Mol Mutagen 39: 82-88.

3. Wolff MS, Collman GW, Barrett C, Huff J (1996) Breast cancer and environmental risk factors: epidemiologic and experimental findings. Annu Rev Pharmacol Toxicol 36: 573-596.

4. Laden F, Hunter DJ (1998) Environmental risk factors and female breast cancer. Annu Rev Public Health 19: 101-123.

5. Kuh D, Ben-Shlomo Y (1997) A life course approach to chronic disease epidemiology: tracing the origins of ill-health from early to later life. Oxford: Oxford University Press.

6. Barker DJP (1992) Fetal and infant origins of adult disease. London: BMJ Publishing.

7. Colditz GA, Frazier AL (1995) Models of breast cancer show that risk is set by events of early life: prevention efforts must shift focus. Cancer Epidemiol Biomarkers Prev 4: 567-571.

8. Trichopoulos D (1990) Hypothesis: does breast cancer originate in-utero? Lancet 335: 939-940.

9. Potischman N, Troisi R (1999) In-utero and early life exposures in relation to risk of breast cancer. Cancer Causes Control 10: 561-573.

10. Lumey LH (1998) Prenatal oestrogens and breast cancer. Paediatr Perinat Epidemiol 12: 361-365.

11. Elliott P, Wakefield JC, Best NG, Briggs DJ (2000) Spatial epidemiology: methods and applications. Oxford: Oxford University Press.

12. GDT (2001) GDT Dynamp/2000 version 11.1, Geographic Data Technology, INC. USA.

13. Ripley BD (1981) Spatial statistics. New York: Wiley.

14. Ripley BD (1977) Modelling spatial patterns. J R Stat Soc Ser B 39: 172-212.

15. Rowlingson B, Diggle PJ (1993) SPLANCS: spatial point pattern analysis code in S-plus. Comput Geosci 19: 627-655.

16. Diggle PJ, Chetwynd AD (1991) Second-order analysis of spatial clustering for inhomogeneous populations. Biometrics 47: 1155-1163.

17. Kulldorff M, Nagarwalla N (1995) Spatial disease clusters: detection and inference. Stat Med 14: 799-810.

18. Kulldorff M (1997) A spatial scan statistic. Commun Stat-Theor M 26:1481-1496.

19. Bailey TC, Gatrell AC (1995) Interactive spatial data analysis. Harlow: Longman.

20. Timander LM, McLafferty S (1998) Breast cancer in West Islip, NY: a spatial clustering analysis with covariates. Soc Sci Med 46: 1623-1636.

21. Rogerson PA, Han D (2002) The effects of migration on the detection of geographic differences in disease risk. Soc Sci Med 55: 1817-1828.

22. McPherson K, Steel CM, Dixon JM (2000) ABC of breast diseases: breast cancer epidemiology, risk factors, and genetics. Br Med J 321: 624-628.

23. Stavola B, Hardy R, Kuh D, Sllva I, Wadsworth M, Swerdlow, A (2000) Birthweight, childhood growth and risk of breast cancer in a British cohort. Br J Cancer 83: 964-968.

24. Cressie N (1993) Statistics for spatial data. New York: Wiley.

25. Reader S (2000) Using survival analysis to study spatial point patterns in geographical epidemiology. Soc Sci Med 50: 985-1000.

26. Snow J (1855) On the mode of communication of cholera. In: Snow on cholera being a reprint of two papers by John Snow, M.D. 1936. New York: The Commonwealth Fund.

27. Jones K, Moon G (1987) Health, disease, and society. London: Routledge.

28. Meade MS, Florin JW, Gesler WM (1988) Medical geography. New York: Guilford.

29. Dunn C, Kingham S, Rowlingson B, Bhopal R, Cockings C, Foy C et al. (2001) Analysing spatially referenced public health data: a comparison of three methodological approaches. Health Place 7: 1-12.

Table 1. Residential history of breast cancer cases and controls: numbers and percentage of complete and missing residences in Erie and Niagara counties: WEB Study, 1996-2001.

	Complete Residence		Incomplete or Missing Residence		Total eligible Erie and Niagara county residence at each time period	
	Case	Control	Case	Control	Case	Control
Birth	505	804	127	189	632	993
	(79.9%)	(81%)	(20.1%)	(19%)		
Menarche	673	1143	98	154	771	1297
	(87.3%)	(88.1%)	(12.7%)	(11.9%)		
First birth	616	1153	97	167	713	1320
	(86.4%)	(87.3%)	(13.6%)	(12.7%)		

Table 2. Characteristics of subjects included in the analysis, subjects with missing residential information, and subjects excluded due to residence outside of the study area (Mean±SD): WEB Study, 1996-2001.

	Cases (*n*=1166)			Controls (*n*=2105)		
Birth	Included (*n*=505)	Missing (*n*=127)	Ineligible* (*n*=534)	Included (*n*=804)	Missing (*n*=189)	Ineligible* (*n*=1112)
Age (years)	56.5±10.9	60.0±11.0	58.9±11.3	55.6±11.7	58.0±11.8	59.4±11.7
Education (years)	13.5±2.4	13.1±2.5	13.6±2.7	13.4±2.2	13.2±2.2	13.3±2.5
Parity	2.2±1.5	2.4±1.7	2.4±1.8	2.6±1.8	2.7±1.8	2.8±1.8
Age at menarche (years)	12.4±1.5	12.6±1.5	12.7±1.7	12.7±1.7	12.6±1.6	12.7±1.7
Age at first birth (years)	24.3±4.6	23.5±4.5	24.2±5.1	24.5±4.3	23.5±4.2	24.0±4.7
Premenopausal (%)	35.2	18.9	26.4	31.7	28.6	24.6
Body Mass Index	28.2±6.4	28.4±5.8	28.7±6.4	28.0±6.2	28.2±6.0	28.4±6.4
Family history of breast cancer (% yes)	21.3	18.9	20.2	12.7	16.2	12.4
History of benign breast disease (% yes)	34.9	37.0	32.8	22.3	25.9	20.6
Menarche	Included (n=673)	Missing (*n*=98)	Ineligible* (*n*=395)	Included (*n*=1143)	Missing (*n*=154)	Ineligible* (*n*=808)
Age (years)	56.6±10.7	60.1±11.6	59.5±11.3	56.0±11.7	60.2±11.7	59.9±11.6
Education (years)	13.5±2.4	12.8±2.6	13.6±2.8	13.4±2.2	13.0±2.3	13.3±2.6
Parity	2.2±1.6	2.8±1.8	2.5±1.8	2.6±1.8	2.9±2.1	2.9±1.8
Age at menarche (years)	12.5±1.6	12.8±1.5	12.7±1.7	12.7±1.6	12.6±1.7	12.7±1.7
Age at first birth (years)	24.3±4.6	23.0±4.3	24.2±5.3	24.4±4.5	23.8±4.4	24.0±4.6
Premenopausal (%)	30.3	24.5	24.6	33.8	23.4	23.3
Body Mass Index	28.1±6.2	29.5±6.4	28.7±6.5	28.3±6.5	27.6±5.5	28.2±6.1
Family history of breast cancer (% yes)	20.2	22.4	20.6	13.1	13.2	12.1
History of benign breast disease (% yes)	34.5	40.8	31.9	22.3	19.5	21.3
First Birth	Included (*n*=616)	Missing (*n*=97)	Ineligible* (*n*=453)	Included (*n*=1153)	Missing (*n*=167)	Ineligible* (*n*=785)
Age (years)	57.4±11.1	58.9±10.8	58.5±11.2	57.0±11.7	60.6±10.7	58.5±12.0
Education (years)	13.4±2.3	13.0±2.9	13.7±2.8	13.3±2.2	13.0±2.1	13.4±2.6
Parity	2.7±1.3	3.1±1.5	1.7±1.9	3.0±1.5	3.4±1.7	2.2±2.0
Age at menarche (years)	12.6±1.5	12.5±1.8	12.6±1.6	12.7±1.6	12.6±1.5	12.7±1.7
Age at first birth (years)	24.8±4.8	22.2±4.1	23.4±4.9	24.7±4.6	22.9±3.5	23.3±4.4
Premenopausal (%)	29.4	26.8	26.0	32.2	18.0	26.6
Body Mass Index	28.4±6.3	30.1±6.6	28.2±6.3	28.1±6.1	28.3±6.5	28.4±6.4
Family history of breast cancer (% yes)	21.2	23.7	18.6	11.5	19.4	13.6
History of benign breast disease (% yes)	34.7	37.1	32.7	21.0	25.1	22.0

* Ineligible due to residence outside of Erie and Niagara county

29

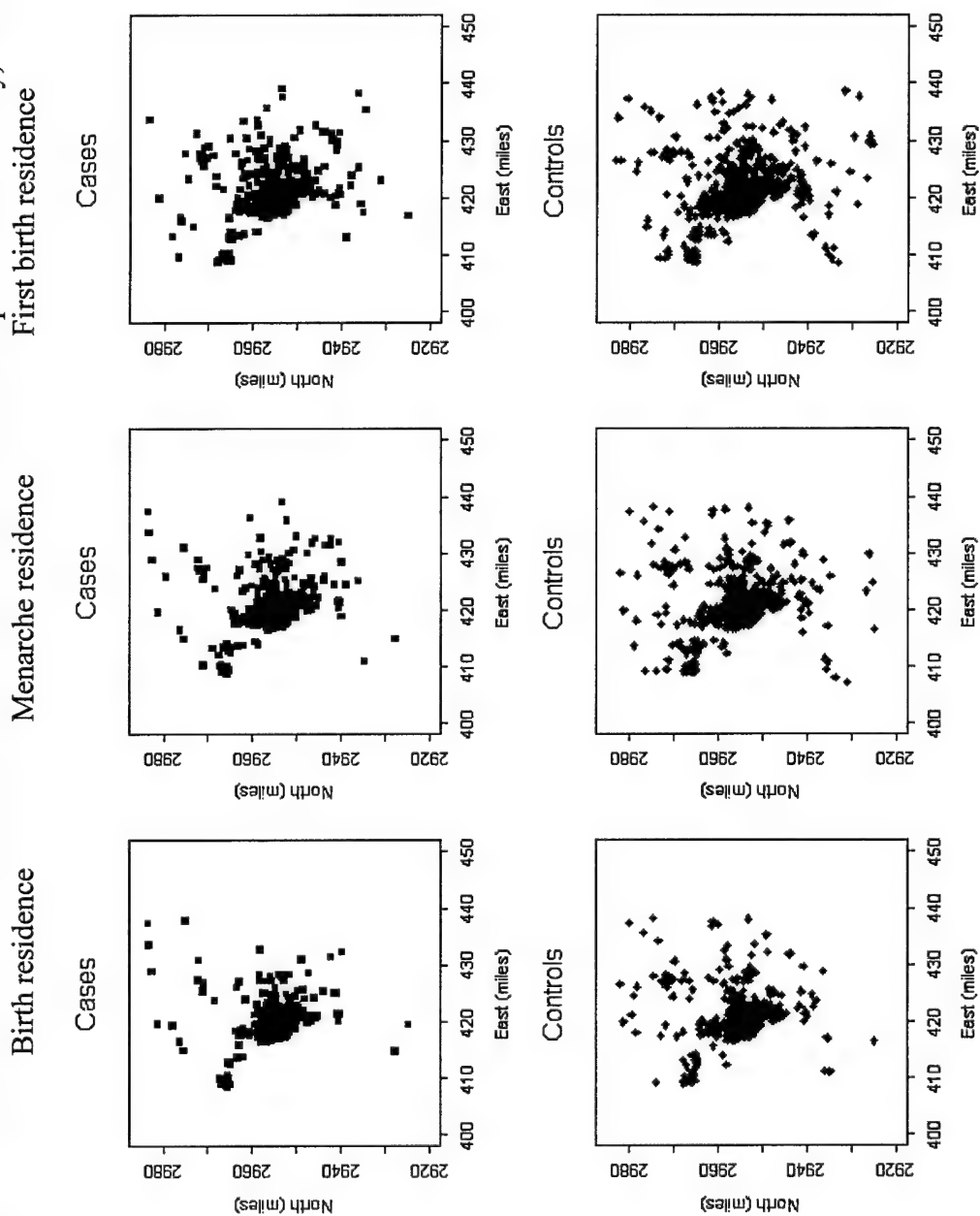Figure 1. Residential location of breast cancer cases and controls at each time period: WEB Study, 1996-2001.

Figure 2. *k*-function differences in clustering patterns between breast cancer cases and controls, WEB Study, 1996-2001:

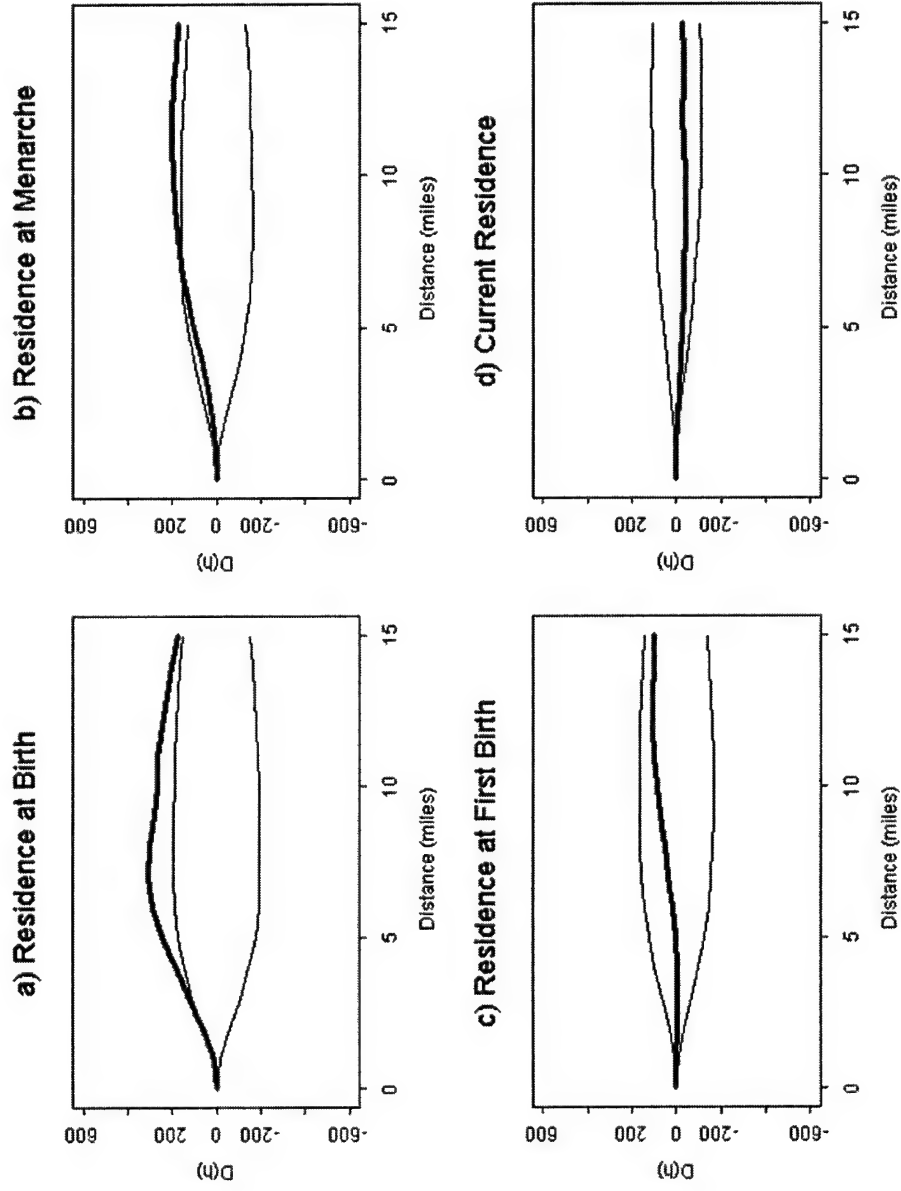Shown are k-function difference in black and 95% confidence limits in grey.

Figure 3. *k*-function differences in clustering patterns between breast cancer cases and controls by menopausal status, WEB Study, 1996-2001.
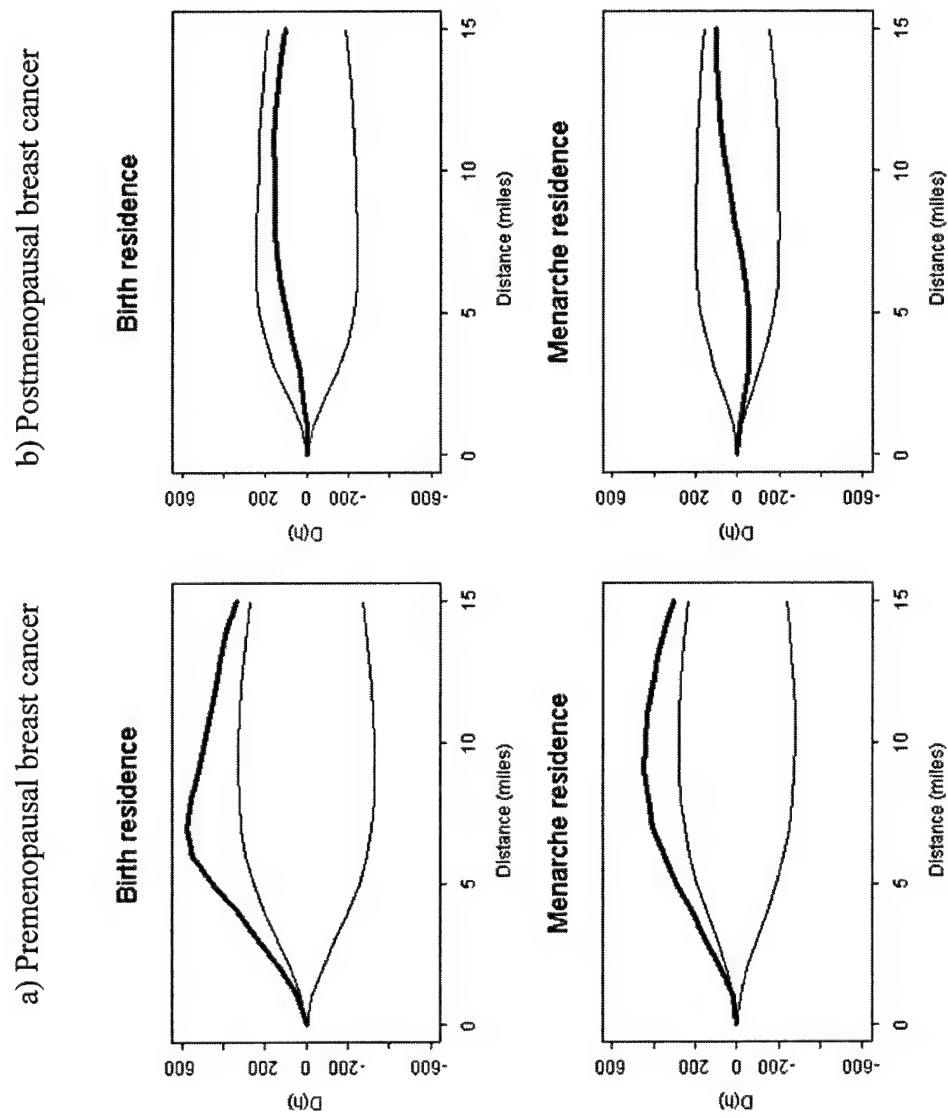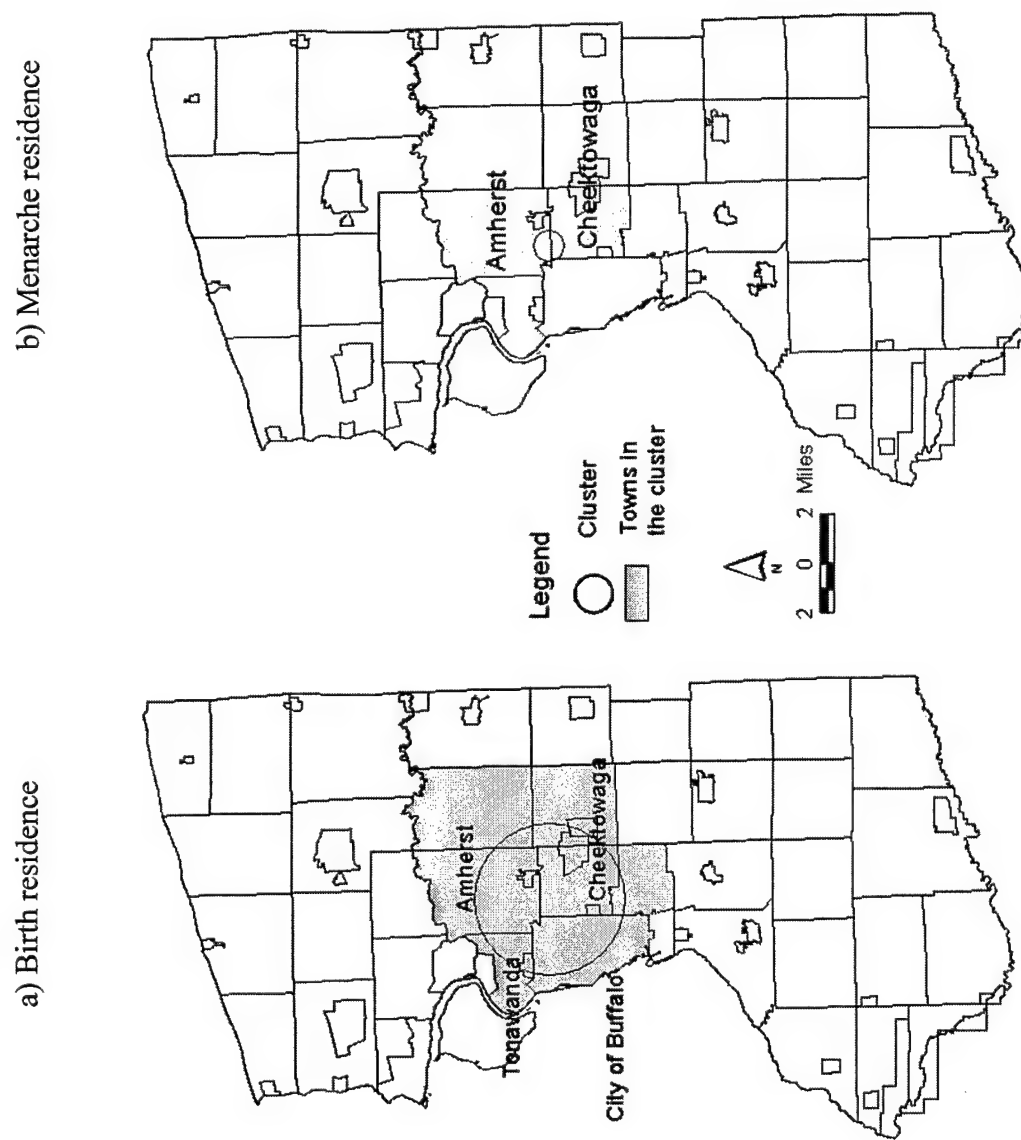
a) Premenopausal breast cancer   b) Postmenopausal breast cancer

32

Figure 4. Geographic clustering of residence at birth and menarche: premenopausal breast cancer, WEB Study, 1996-2001

a) Birth residence

b) Menarche residence

33

APPENDIX III

Breast Cancer Risk and Exposure in Early Life to Polycyclic Aromatic Hydrocarbons Using Total Suspended Particulates as a Proxy Measure

Matthew R. Bonner[1], Daikwon Han[1], Jing Nie[1], Peter Rogerson[2], John E. Vena[3], Paola Muti[1], Maurizio Trevisan[1], Stephen Edge[4], Jo L. Freudenheim[1]

[1] Department of Social and Preventive Medicine, School of Public Health and Health Professions, University at Buffalo, Buffalo, New York 14214.

[2] Department of Geography, University at Buffalo, Buffalo, New York 14214.

[3] Department of Epidemiology and Biostatistics, Arnold School of Public Health, University of South Carolina, Columbia, South Carolina 29208.

[4] Department of Surgery, Roswell Park Cancer Institute, Buffalo, NY 14263.

Corresponding author: Matthew R. Bonner, Occupational and Environmental Epidemiology Branch, National Cancer Institute, 6120 Executive Blvd. EPS 8121, MSC 7240, Bethesda, MD 20892-7240 Phone: (301) 402-7825, Fax: (301) 4021819, E-mail address: bonnerm@mail.nih.gov

Running title: Polycyclic aromatic hydrocarbons and breast cancer

Abstract

Polycyclic aromatic hydrocarbons (PAHs) are ubiquitous in the environment. We hypothesized that early life exposure to PAHs may have particular importance in the etiology of breast cancer. We conducted a population-based, case-control study of ambient PAH exposure in early life in relation to the risk of breast cancer. Total suspended particulates (TSP), a measure of ambient air pollution, was used as a proxy for PAH exposure. Cases (n=1,166) were women with histologically-confirmed, primary, incident breast cancer. Controls (*n*=2,105) were frequency matched by age, race, and county of residence to cases. Annual average TSP concentrations (1959-1997) by location were obtained from the New York State Department of Environmental Conservation for Erie and Niagara Counties. Based on the monitor readings, prediction maps of TSP concentrations were generated with ArcGIS 8.0 (ESRI, Inc., Redlands, CA) using inverse distance squared weighted interpolation. Unconditional logistic regression was used to estimate odds ratios (OR) and 95% confidence intervals (95% CI). In postmenopausal women, exposure to high concentrations of TSP (>140 $\mu g/m^3$) was associated with an adjusted OR of 2.42 (95% CI=0.97-6.09) compared with exposure to low concentrations (<84 $\mu g/m^3$). However, in premenopausal women, where exposures were generally lower, the results were inconsistent with our hypothesis and in some instances were suggestive of a reduction in the risk of breast cancer. Our study suggests that exposure in early life to high levels of PAHs may increase the risk of postmenopausal breast cancer; however other confounders related to geography cannot be ruled out.

Introduction

Polycyclic aromatic hydrocarbons (PAHs) are ubiquitous in the environment and commonly present in particulate air pollution (*1, 2*). PAHs are a broad category of chemical compounds composed of carbon and hydrogen and are formed as a by-product during combustion of organic materials. Important sources of PAHs include cigarette smoke, steel mills, foundries, automobiles, coal combustion for electricity production and many other industrial and non-industrial processes. PAHs are also found in food and are formed when food is cooked at high temperatures (i.e., grilling meats). In addition to anthropogenic sources, natural sources (i.e., volcanoes and forest fires) also contribute PAHs to the atmosphere (*1, 3*). Most of these sources not only contribute to the release of PAHs into the environment, but also contribute to particulate air pollution. Ninety to 95% of particulate phase PAHs are physically associated with particulate matter less than 3.3 μm (*2, 4*). These small particles are thought to have particular biologic relevance since they can be inhaled and deposited in the lower respiratory tract.(*5*) PAHs are lipophilic (*6, 7*) They have been shown to be mammary carcinogens in animal models (*1, 8, 9*). and there is evidence that they may also be human mammary carcinogens (*10, 11*). In addition, PAHs may also have estrogenic and antiestrogenic properties that could potentially affect breast cancer risk (*12*).

No studies have examined exposure to total suspended particulates and breast cancer risk and only a few epidemiologic investigations of breast cancer have examined PAHs. Petralia and colleagues (*13*) examined premenopausal breast cancer and occupational exposure to benzene and PAHs using job-exposure matrices in a population-based, case-control study. High probability of occupational exposure to benzene and PAHs was

associated with premenopausal breast cancer. However, because women were exposed to a mixture of compounds, it was difficult in that study to estimate whether PAHs had an independent effect on breast cancer risk.

Rundle et al.(*11*) examined PAH-DNA adducts in breast tumor tissue. They found a 2-fold increase in PAH-DNA adducts in malignant tumors compared with tissue from controls with benign breast disease with atypia. Gammon et al.(*10*) examined PAH-DNA adducts in mononuclear cells in relation to the risk of breast cancer in a case-control study of Long Island residents. They found a nearly 50% increase in the risk of breast cancer for subjects in the highest quintile of PAH-DNA adducts in mononuclear cells; there was no dose-response relationship.

Early life exposures, including exposure to PAHs, may have particular importance in the etiology of breast cancer (*14*). Early age at exposure to ionizing radiation, for example, confers increased risk of breast cancer when compared with later age at exposure (*15, 16*). In addition, several other established risk factors also indicate the importance of early life factors in the etiology of breast cancer. Breast cancer risk is increased in women with earlier age at menarche, whereas earlier age at first birth reduces the risk of breast cancer. The physiological changes that occur to breast tissue during development further support the postulation that early life exposures may be important. Around menarche, the mammary gland begins to develop and differentiate into defined ducts and lobules. The primary lobules formed at this time are type 1 lobules. These lobules further differentiate into type 2 and type 3 lobules during pregnancy (*17*). *In vitro* studies have shown that cells from type 1 lobules are more sensitive to proliferation signals than either cells from type 2 or 3 lobules (*18*). In

addition, human breast epithelial cells from type 1 lobules were more sensitive to the transforming effects of the PAH, 7,12-dimethlybenzo(a)anthracene and N-methyl-N-nitrosourea than were type 3 lobule cells (*19*).

We conducted a population-based, case-control study of PAH exposure in early life in relation to the risk of breast cancer using total suspended particulates (TSP), a measure of ambient air pollution, as a proxy for PAH exposure. We examined time periods that are thought to be critical exposure periods with regards to susceptibility to breast cancer: at the time of birth, at menarche, and at the time of when the participant first gave birth.

Materials and Methods

The Western New York Exposures and Breast Cancer Study (WEB Study) is a population-based, case-control study was conducted with women living in Erie and Niagara Counties in Western New York State during 1996-2001. All participants were aged 35-79 years. Cases included 1,166 women with histologically-confirmed, primary, incident breast cancer. In addition, cases under the age of 65 years were restricted to women with a driver's license. Controls (*n*=2,105) were frequency matched by age, race, and county of residence to cases. Controls under the age of 65 years were randomly selected from the New York State Department of Motor Vehicles driver's license list and controls 65 years of age and over were randomly selected from the Healthcare Financing Administration Medicare rolls. For these analyses, cases and controls were restricted to participants who were residents of Erie and Niagara Counties during each of the three pertinent time periods; birth, menarche, and first birth. A total of 1,638 cases and 3,396 controls met our inclusion criteria of between 35-79 years of age, current resident of Erie or Niagara County, no previous cancer diagnosis other than non-melanoma skin cancer

and an ability to speak English. The response rates were 71% (1,166/1,638) and 62% (2,105/3,396) for cases and controls, respectively. All participants provided informed consent; the protocol was approved by the Institutional Review Boards of the University at Buffalo School of Medicine and Biomedical Sciences and of participating hospitals.

Data Collection

Using extensive in-person interviews and self-administered questionnaires, participants provided information regarding medical history, diet, alcohol consumption, smoking history, lifetime passive smoke exposure, occupational history, and residential history. Residential histories were reported by the subject dating back to birth. For addresses in Erie and Niagara Counties, Polk and city directories were searched to find missing address information. For addresses with missing zip codes, we used ZP4 (Semaphore Corporation, Aptos, CA), a commercially available database that uses information about street name and number and city designation to find missing zip codes. Residential histories and interview data were used to identify each subject's residence at her birth, menarche, and her first birth. These addresses were geocoded with ArcView 3.2 (ESRI, Inc., Redlands, CA) using Dynamap 2000 (GDT Inc., Lebanon, NH) as the reference theme (i.e., street map) of Erie and Niagara Counties.

Exposure Assessment

The New York State Department of Environmental Conservation maintains air monitors that began measuring total suspended particulates (TSP) in 1959. These monitors measured TSP concentrations every seven days. Annual average TSP concentrations (1959-1997) were obtained from these monitors for Erie and Niagara

Counties. In total, 87 monitors were operating at various times in Erie and Niagara Counties. For the period of the 1960's, there were fewer monitors operating than at later time periods. There was very little within monitor variation of TSP concentration during this time period and average TSP concentrations were calculated for the entire decade for each monitor. By averaging the TSP concentrations for each monitor, the overall TSP estimates were more stable. Considerably more monitors were operating in the years after 1969. Annual average TSP concentrations were calculated for each year for 1970 through 1997 for each monitor. In addition to TSP, ambient benzo(a)pyrene (B(a)P) was measured between November 1, 1973 and November 1, 1974 in Erie County, New York for 11 of the 87 monitoring sites. The Pearson correlation coefficient between the measured log transformed TSP and log transformed B(a)P concentrations at these 11 monitoring sites was 0.90, suggesting that the ambient TSP concentrations reasonably estimate ambient PAHs concentrations in this region.

Based on the monitor readings for each time period, prediction maps of TSP concentrations were generated with ArcGIS 8.0 (ESRI, Inc., Redlands, CA) using inverse distance squared weighted interpolation. We assumed a 45-degree angle to account for the prevailing southwesterly winds and limited the exposure estimation for each address to the seven closest sampling monitors. The primary assumption of these geostatistical methods is that close locations are more similar to one another than are locations relatively farther away.(20) The estimated individual residential TSP concentrations were insensitive to changing the number of monitors included for the exposure estimation. In total, 29 prediction maps were constructed of estimated TSP concentrations for the two-county region; one for the 1960's and one for each year after

that until 1997. These maps were used to determine exposure to TSP at each participant's address for the relevant time period. The 1960's TSP concentration prediction map is provided as an example in figure 1.

TSP concentrations for addresses before the 1960s were estimated assuming that the interpolated concentrations in the 1960s were representative of earlier time periods. Industrialization in Erie and Niagara Counties began at the end of the 19th century and the industrial activities that contributed most heavily to air pollution were very active prior to the 1960's and was relatively constant over the time period (21). Further, measures to control air quality were not implemented until the early 1970's. Consequently, the 1960's concentrations of TSP probably reflect ambient levels in the earlier time period.

Statistical Analysis

Unconditional logistic regression (22) was used to estimate odds ratios (OR) and 95% confidence intervals (95% CI). TSP concentrations were categorized into 4 levels (<84 ug/m^3, 84-114 ug/m^3, 115-140ug/m^3, and >140ug/m^3). The cut points for the categorical analyses were derived from the quartiles of the distribution of measurements of TSP concentrations in the 1960s. In addition to the categorical analysis, we examined TSP concentrations on a continuous scale. Further, logistic quadratic spline regression with knots at 84 μg/m^3 and 140 μg/m^3 was used to graphically depict the exposure-response trend; the estimated probability of being a case was calculated from the quadratic spline regression equation and adjusted for age, education and parity. The values for the two knots in the spline regression were selected based on the previous categorical analysis. The end categories were restricted to linear segments to prevent instability (23).

We considered age, race, education, age at first birth, age at menarche, parity, previous benign breast disease, family history of breast cancer, body mass index (weight (kg)/height (m)^2), and age at menopause as potential confounders in multivariate logistic regression. The models presented include age, education, and parity and was determined by excluding variables from the full model which did not alter the risk estimates more than 10%. All models were stratified by menopausal status. *P* for trend statistics were determined by the *p*-value for the coefficient of the continuous exposure variable, while adjusting for covariates.

Results

The descriptive characteristics of the participants included in the birth, menarche, first birth, and the overall case-control study are depicted in Table 1. There were no major differences between the distributions of these variables between each time period. We were able to successfully geocoded 79%, 87%, and 87% of the Erie and Niagara County birth, menarche, and first birth addresses, respectively.

Exposure to concentrations of TSP >84 μg/m^3 at the time of birth was associated with an increase in the odds ratio for premenopausal women (Table 2), however, there was no exposure-response relationship and the *P* for trend was not significant. In addition, there were relatively few participants exposed to the lowest concentrations of TSP and this resulted in wide confidence intervals for the corresponding point estimates. In postmenopausal women, exposure to high concentrations of TSP (>140 μg/m^3) was associated with an adjusted OR of 2.42 (95% CI=0.97-6.09) compared with exposure to low concentrations (<84 μg/m^3). For risk associated with estimated residential TSP concentrations on a continuous scale, in postmenopausal women, we observed a 21%

increase in the odds ratio for every 30 $\mu g/m^3$ increase in TSP concentration (adjusted OR = 1.20, 95% CI = 1.04-1.38). In the spline regression analysis, there was an increase in the probability of being a case with an increase in TSP concentration (Figure 2). No increase in risk was observed for premenopausal women on a continuous scale (OR = 0.92, 95% CI = 0.76-1.11 for every increase in 30$\mu g/m^3$ of TSP). Further, the spline regression analysis for the premenopausal women indicated an inverted parabola exposure-response relationship with increasing TSP concentration (Figure 3).

At menarche, exposure to high concentrations of TSP was also associated with a modest increase in the odds ratio for postmenopausal women with exposure > 84 $\mu g/m^3$, although the *P* for trend was not significant (Table 3). In the continuous analysis, for every 30 $\mu g/m^3$ increase in TSP concentrations, the odds ratio increased 8% (adjusted OR = 1.08, 95% CI = 0.96-1.21) for postmenopausal women. The risk estimates for the premenopausal women were not consistent with our hypothesis. In this group, there was a non-significant reduction in risk in the highest exposure category (adjusted OR = 0.66, 95% CI = 0.38-1.16). Exposure to high concentrations of TSP at the time of first birth was also associated with a modest increase in the odds ratio for postmenopausal women (Table 4). For premenopausal women exposed to high concentration of TSP, there was some indication of a non-significant reduction in the odds ratio (OR = 0.52, 95% CI = 0.22-1.20).

Discussion

While numerous epidemiologic studies have investigated the carcinogenicity of air pollution in relation to lung cancer, (*24-26*) to our knowledge, no investigations have examined exposure to total suspended particulates and breast cancer. The findings from

this study suggest that early life exposure to high concentrations of TSP, a proxy measure of PAHs, may be associated with an increased risk of breast cancer in postmenopausal women. We found more than a two-fold increase in risk for those with a birth residence in areas where exposure was greater than 140 $\mu g/m^3$ compared with those with a birth residence where concentration was less than 84 $\mu g/m^3$. Exposure at menarche and first birth were less strongly associated with risk. There was little evidence that exposure to high concentrations of TSP was positively associated with premenopausal breast cancer. However, the inconsistency of these findings for women in this group may be attributable to insufficient induction time between exposure in early life and the occurrence of breast cancer and to secular changes in exposure levels.

Several previous studies have examined PAH exposure in adult life in relation to cancer (*10, 11*). There is evidence that PAH-DNA adducts in tumor tissue and peripheral blood tend to be higher in breast cancer cases that in controls. Tumor PAH-DNA adducts levels are markers of recent exposure and PAH-DNA adducts in mononuclear cells are at best indicative of exposure several years prior to collection. Our findings are based on historical estimates of early life exposure. They support the hypothesis that PAH exposure may be associated with breast cancer risk and indicate that early life exposure to these compounds may have particular relevance to the etiology of breast cancer.

Other exposures, particularly ionizing radiation, have been observed to increase risk of breast cancer with early age at exposure. Similarly, exposure to PAHs in early life may also confer increased risk of breast cancer compared with adult exposure to PAHs. In addition, there is some evidence that early life exposure to PAHs could impact the developing fetus. In a study of early life exposure to high PAHs concentrations in air,

Perera and colleagues found PAH exposure to be associated with reduced birth weight, birth length, and head circumference (*27*). Several studies investigating the relationship between birth weight and the risk of breast cancer have observed a j-shaped curve with birth weight: Those <2,500 g at birth had increased risk of breast cancer compared with women with birth weights of 2500-2999g (*28, 29*).

It is also possible that PAHs may not affect breast cancer risk and our findings are a result of other carcinogens and co-carcinogens found in total suspended particulates. We speculated that PAHs physically associated with TSP may be the agent responsible for the association between TSP and breast cancer risk that we observed. However, we cannot rule out the possibility that other compounds present in TSP are affecting breast cancer risk or are acting synergistically with PAHs. In experimental studies, for instance, application of coal tar produced more skin tumors than did the application of only benzo(a)pyrene, which is thought to be the primary carcinogen in coal tar. Other constituents in coal tar seem to contribute to the carcinogenic potential and enhance synergistically the effect of benzo(a)pyrene (*9*). It may be that it is the mixture of compounds in TSP that is relevant to breast cancer risk.

Several methodological concerns need to be considered when interpreting our findings. Foremost is the potential for selection bias to affect the internal validity of the study. To investigate the extent of the geographic selection bias, we compared the geographic distribution of breast cancer cases in the study with that of the breast cancer cases reported to the New York State Tumor Registry. The expected number of cases per zip code in Erie and Niagara Counties were obtained from the NY State Tumor Registry and compared with the number of cases identified for our study. Overall, there was some

evidence that cases identified for this study tended to reside more closely to the study site than cases identified in the NY State Tumor Registry. When the expected number of controls per zip code (obtained from the 1990 U.S. Census) was compared with the observed number of controls, controls were also more likely to currently reside more closely to the study site.

In addition, there is the possibility that our results were biased because the sample was restricted to women who were both current residents of Erie or Niagara Counties at the time of the case-control study and who had lived there during their earlier life. However, we found little difference between those subjects with birth addresses in Erie and Niagara Counties compared with those subjects with birth addresses outside of these two counties with regards to demographic characteristics or established risk factors (data not shown).

Small numbers in some categories and the resultant large confidence intervals affected our ability to draw conclusions from our data. The distribution of TSP concentrations contributed to the small numbers in certain categories. Ambient TSP concentrations had large spatial variation in the 1960s, but in general, TSP concentrations were high compared with later time periods. However, TSP concentrations began to decrease in the early 1970's leading to low estimates in the 1970s-90's with very little geographic variation in TSP concentrations. Consequently, the distributions for each time period were very different. Few postmenopausal participants were exposed to low concentrations at birth and few premenopausal women were exposed to high concentrations at the time of first birth. These trends in ambient air concentration of TSP precluded an analysis of exposure in adult life up to the time of diagnosis because the lack of variability. In order to be able to make comparisons between time periods, we

chose to use a common cut point for all analyses. The cut points for our analyses were arbitrarily selected based on the distribution of the TSP measurements in the 1960s. With the majority of participants having had high levels of TSP at birth, these cut points resulted in small numbers in the referent group. However, the continuous and spline regression analyses support the direction of the association in postmenopausal women.

In addition to the secular changes in ambient TSP concentrations, TSP is a relatively crude measure of ambient air pollution. In 1987, it was replaced with particulate matter <10 microns (30). Currently, particulate matter <2.5 microns is considered to be the most relevant measure for biologic effects of air pollution because these fine particles are respired into the lower respiratory tract (5). However, TSP concentrations were the only consistently measured ambient air pollutant in the early 1960's, the period before the Clean Air Act, which led to reductions in ambient air pollution. TSP is the best available measure to estimate historical exposure to air pollution. Nevertheless, there remains the potential for exposure misclassification because TSP concentration measurements were used as a surrogate for exposure to PAHs. PAHs exist in the ambient air in both the gaseous and particulate phase. The use of TSP captures exposure to PAHs in the particulate phase only (31), although ambient B(a)P concentrations were highly correlated (r = 0.90) with TSP concentrations in this region. In addition, the interpolation method used to estimate concentrations of TSP at residential addresses likely contributed some error. The air samplers were not randomly distributed throughout Erie and Niagara Counties. In general, air samplers were placed in regions thought to have high levels of air pollution. Because the monitoring system was not designed to provide county wide

characterization of TSP levels, some outlying areas were never monitored and were approximately 18 miles from the closest monitor.

Another potential problem in assessing exposure arose because humans are peripatetic (*32*). Therefore, our estimates of TSP concentrations are site specific for each participant and may not represent exposures at other places where these participants spent time. This is likely less of a problem for the analyses of birth residence. By menarche, however, these participants would spend a considerable proportion of their time away from home.

In summary, we examined exposure to total suspended particulates, a surrogate for PAHs exposure, in relation to the risk of breast cancer. We found a suggestion of an association between exposure to high concentration of TSP at birth and an increase risk of breast cancer in postmenopausal women. Among premenopausal women, there was no evidence of such an association with risk of breast cancer. While, these results are suggestive, they necessarily should be considered preliminary. Future research on the effects of early life exposure to PAHs and other related compounds is warranted.

References

1.   Bostrom, C. E., Gerde, P., Hanberg, A., Jernstrom, B., Johansson, C., Kyrklund, T., Rannug, A., Tornqvist, M., Victorin, K., and Westerholm, R. Cancer risk assessment, indicators, and guidelines for polycyclic aromatic hydrocarbons in the ambient air. Environ Health Perspect, *110 Suppl 3:* 451-88, 2002.

2.   Ravindra, Mittal, A. K., and Van Grieken, R. Health risk assessment of urban suspended particulate matter with special reference to polycyclic aromatic hydrocarbons: a review. Rev Environ Health, *16:* 169-89, 2001.

3.   EPA, U. S. Locating and Estimating Air Emissions from Sources of Polycyclic Organic Matter. Research Triangle Park, NC: United States Environmental Protection Agency, 1998.

4.   Perera, F. P., and Weinstein, I. B. Molecular epidemiology and carcinogen-DNA adduct detection: new approaches to studies of human cancer causation. J Chronic Dis, *35:* 581-600, 1982.

5.   Dockery, D. W., Pope III, C.A. Outdoor Air 1: Particulates. *In:* K. Steenland, & Savitz, D.A. (ed.), Topics in Environmental Epidemiology. New York: Oxford University Press, 1997.

6.   Obana, H., Hori, S., Kashimoto, T., and Kunita, N. Polycyclic aromatic hydrocarbons in human fat and liver. Bull Environ Contam Toxicol, *27:* 23-7, 1981.

7.   Morris, J. J., and Seifter, E. The role of aromatic hydrocarbons in the genesis of breast cancer. Med Hypotheses, *38:* 177-84, 1992.

8.   Phillips, D. H. Fifty years of benzo(a)pyrene. Nature, *303:* 468-72, 1983.

9.   Rubin, H. Synergistic mechanisms in carcinogenesis by polycyclic aromatic hydrocarbons and by tobacco smoke: a bio-historical perspective with updates. Carcinogenesis, *22:* 1903-30, 2001.

10.   Gammon, M. D., Santella, R. M., Neugut, A. I., Eng, S. M., Teitelbaum, S. L., Paykin, A., Levin, B., Terry, M. B., Young, T. L., Wang, L. W., Wang, Q., Britton, J. A., Wolff, M. S., Stellman, S. D., Hatch, M., Kabat, G. C., Senie, R., Garbowski, G., Maffeo, C., Montalvan, P., Berkowitz, G., Kemeny, M., Citron, M., Schnabel, F., Schuss, A., Hajdu, S., and Vinceguerra, V. Environmental toxins and breast cancer on Long Island. I. Polycyclic aromatic hydrocarbon DNA adducts. Cancer Epidemiol Biomarkers Prev, *11:* 677-85, 2002.

11.   Rundle, A., Tang, D., Hibshoosh, H., Estabrook, A., Schnabel, F., Cao, W., Grumet, S., and Perera, F. P. The relationship between genetic damage from polycyclic

aromatic hydrocarbons in breast tissue and breast cancer. Carcinogenesis, *21:* 1281-9, 2000.

12.   Santodonato, J. Review of the estrogenic and antiestrogenic activity of polycyclic aromatic hydrocarbons: relationship to carcinogenicity. Chemosphere, *34:* 835-48, 1997.

13.   Petralia, S. A., Vena, J. E., Freudenheim, J. L., Dosemeci, M., Michalek, A., Goldberg, M. S., Brasure, J., and Graham, S. Risk of premenopausal breast cancer in association with occupational exposure to polycyclic aromatic hydrocarbons and benzene. Scand J Work Environ Health, *25:* 215-21, 1999.

14.   Potischman, N., and Troisi, R. In-utero and early life exposures in relation to risk of breast cancer. Cancer Causes Control, *10:* 561-73, 1999.

15.   Land, C. E. Studies of cancer and radiation dose among atomic bomb survivors. The example of breast cancer. Jama, *274:* 402-7, 1995.

16.   Tokunaga, M., Land, C. E., Tokuoka, S., Nishimori, I., Soda, M., and Akiba, S. Incidence of female breast cancer among atomic bomb survivors, 1950-1985. Radiat Res, *138:* 209-23, 1994.

17.   Russo, J., Hu, Y. F., Yang, X., and Russo, I. H. Developmental, cellular, and molecular basis of human breast cancer. J Natl Cancer Inst Monogr: 17-37, 2000.

18.   Russo, J., Calaf, G., Sohi, N., Tahin, Q., Zhang, P. L., Alvarado, M. E., Estrada, S., and Russo, I. H. Critical steps in breast carcinogenesis. Ann N Y Acad Sci, *698:* 1-20, 1993.

19.   Russo, J., Reina, D., Frederick, J., and Russo, I. H. Expression of phenotypical changes by human breast epithelial cells treated with carcinogens in vitro. Cancer Res, *48:* 2837-57, 1988.

20.   Johnston, K., Ver Hoef, J.M., Krivoruchko, K., Lucas, N. Using ArcGIS Geostisticical Analyst. Redlands, CA: ESRI, 2001.

21.   Vena, J. E. Air Pollution and Lung Cancer in Erie County, New York. Department of Social and Preventive Medicine. Buffalo, NY: University at Buffalo, 1980.

22.   Hosmer, D. a. L., S. Applied Logistic Regression. New York: Wiley, 2000.

23.   Greenland, S. Dose-response and trend analysis in epidemiology: alternatives to categorical analysis. Epidemiology, *6:* 356-65, 1995.

24.   Pope, C. A., 3rd, Burnett, R. T., Thun, M. J., Calle, E. E., Krewski, D., Ito, K., and Thurston, G. D. Lung cancer, cardiopulmonary mortality, and long-term exposure to fine particulate air pollution. Jama, *287:* 1132-41, 2002.

25.   Vena, J. E. Air pollution as a risk factor in lung cancer. Am J Epidemiol, *116:* 42-56, 1982.

26.   Nyberg, F., Gustavsson, P., Jarup, L., Bellander, T., Berglind, N., Jakobsson, R., and Pershagen, G. Urban air pollution and lung cancer in Stockholm. Epidemiology, *11:* 487-95, 2000.

27.   Perera, F. P., Whyatt, R. M., Jedrychowski, W., Rauh, V., Manchester, D., Santella, R. M., and Ottman, R. Recent developments in molecular epidemiology: A study of the effects of environmental polycyclic aromatic hydrocarbons on birth outcomes in Poland. Am J Epidemiol, *147:* 309-14, 1998.

28.   Titus-Ernstoff, L., Egan, K. M., Newcomb, P. A., Ding, J., Trentham-Dietz, A., Greenberg, E. R., Baron, J. A., Trichopoulos, D., and Willett, W. C. Early life factors in relation to breast cancer risk in postmenopausal women. Cancer Epidemiol Biomarkers Prev, *11:* 207-10, 2002.

29.   Sanderson, M., Williams, M. A., Malone, K. E., Stanford, J. L., Emanuel, I., White, E., and Daling, J. R. Perinatal factors and risk of breast cancer. Epidemiology, *7:* 34-7, 1996.

30.   EPA, U. S. Revisions to the National Ambient Air Quality Standards for Particulate Matter:Final Rules., Code of Federal Regulations 40 CFR Part 50, 1987.

31.   Harrison, R. M. Measurements of Concentrations of Air Pollutants. *In:* S. T. Holgate, Samet, J.M., Koren, H.L., and Maynard, R.L. (ed.), Air Pollution and Health, pp. 69. San Diego: Academic Press, 1999.

32.   Griffith, J. A., T.E. Characterizing Human Exposure. Environmental Epidemiology. New York: Van Nostrand Reinhold, 1993.

Figure 1. Total Suspended Particulate Concentrations in Erie and Niagara Counties,
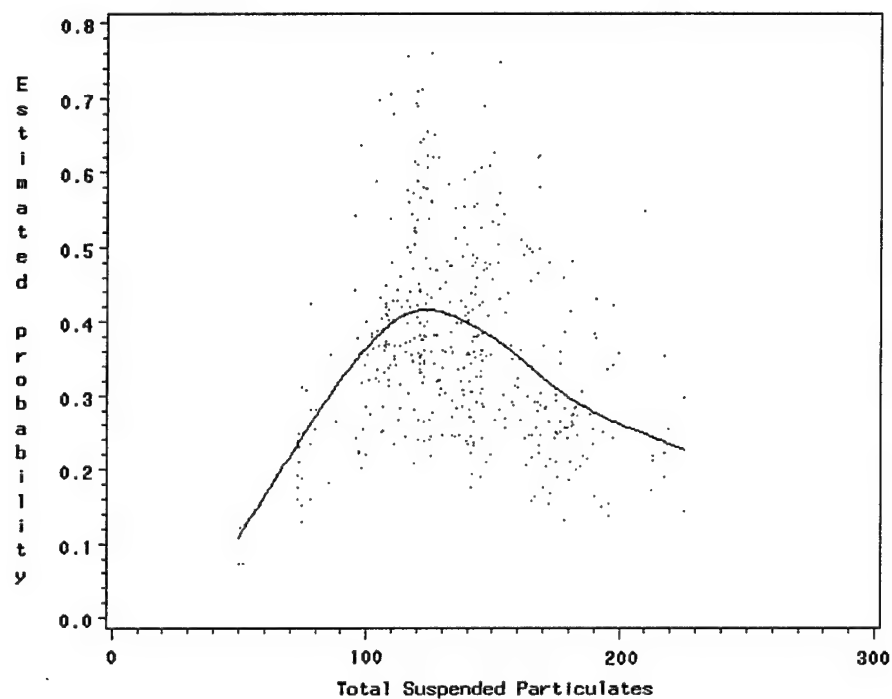
Western New York (1960's).
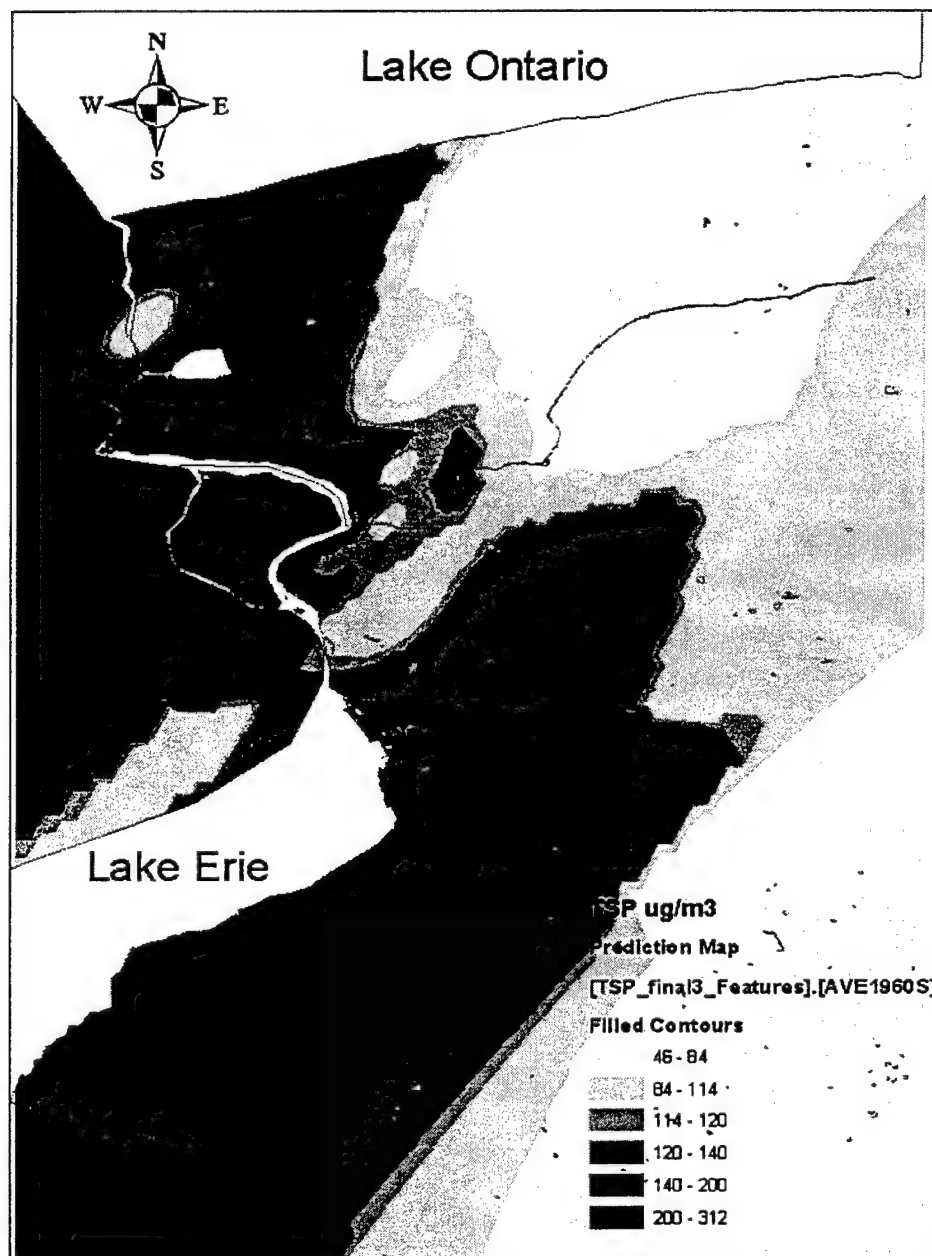
Figure 2. Estimated Probability of Being a Case for Postmenopausal Women by Total

Suspended Particulate Concentration (ug/m^3) at Birth Address.
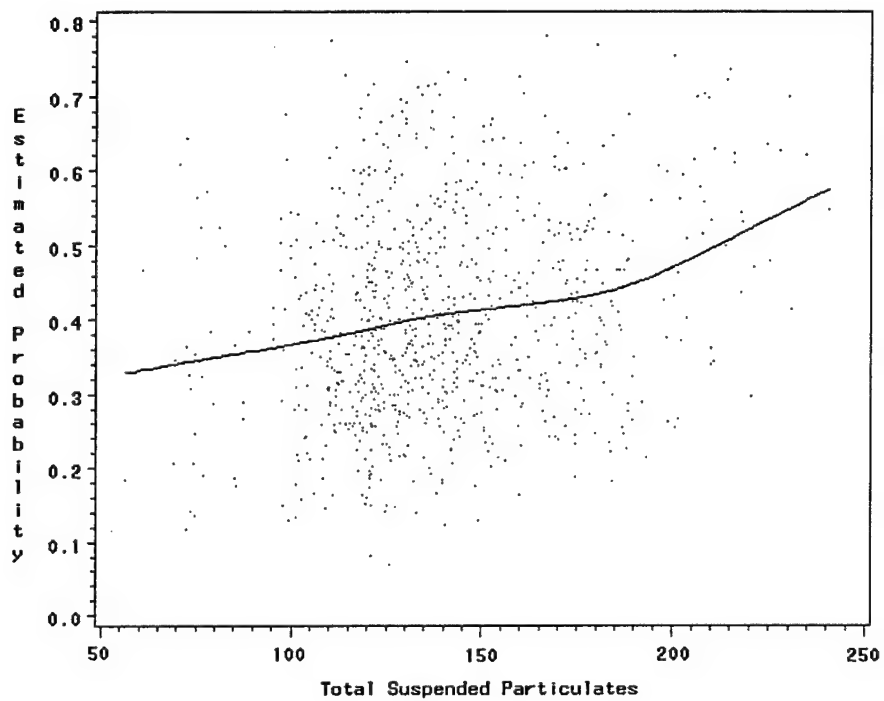
Figure 3. Estimated Probability of Being a Case for Premenopausal Women by Total

Suspended Particulate Concentration (ug/m^3) at Birth Address.

Table 1. Descriptive Characteristics for Study Participants at Birth, Menarche, First Birth, and Overall Study: Western New York Exposures and Breast Cancer Study (WEB Study).

	Premenopausal Women							
	Birth		Menarche		First Birth		Overall Study	
	Cases	Controls	Cases	Controls	Cases	Controls	Cases	Controls
	(n=164)	(n=283)	(n=204)	(n=386)	(n=181)	(n=371)	(n=325)	(n=610)
Age [yrs; mean (SD)]	44.3 (4.5)	43.8 (4.5)	44.5 (4.5)	43.8 (4.6)	44.5 (4.7)	44.2 (4.6)	44.9 (4.6)	44.1 (4.6)
Education [yrs; mean (SD)]	13.7 (2.6)	14.2 (2.2)	13.9 (2.0)	14.1 (2.2)	13.9 (2.0)	14.1 (2.1)	14.0 (2.3)	14.2 (2.2)
Age at Menarche [yrs; mean (SD)]	12.5 (1.5)	12.6 (1.6)	12.5 (1.6)	12.6 (1.6)	12.5 (1.4)	12.7 (1.6)	12.5 (1.6)	12.6 (1.6)
Age at First Birth [yrs; mean (SD)]	25.1 (4.8)	26.1 (4.5)	25.4 (4.9)	25.7 (4.8)	25.7 (7.2)	26.1 (4.9)	25.0 (5.1)	25.8 (4.8)
Body Mass Index [mean; (SD)]	27.4 (7.2)	27.3 (6.3)	27.0 (6.9)	27.6 (6.8)	27.2 (7.2)	27.3 (6.6)	27.2 (6.8)	27.6 (6.7)
Benign Breast Disease (yes)	34%	22%	35%	20%	35%	22%	37%	21%
1° Relative with Breast Cancer (yes)	23%	10%	23%	10%	21%	9%	21%	10%

Table 1, continued. Descriptive Characteristics for Study Participants at Birth, Menarche, First Birth, and Overall Study: Western New York Exposures and Breast Cancer Study (WEB Study).

	Postmenopausal Women							
	Birth		Menarche		First Birth		Overall Study	
	Cases	Controls	Cases	Controls	Cases	Controls	Cases	Controls
	(n=357)	(n=524)	(n=469)	(n=757)	(n=435)	(n=782)	(n=841)	(n=1495)
Age [yrs; mean (SD)]	62.3 (7.9)	62.1 (9.1)	61.9 (8.1)	62.2 (9.1)	62.8 (8.3)	63.0 (8.9)	63.0 (8.5)	63.4 (8.9)
Education [yrs; mean (SD)]	13.4 (2.5)	13.0 (2.1)	13.4 (2.5)	13.0 (2.1)	13.2 (2.4)	13.0 (2.2)	13.3 (2.6)	13.0 (2.3)
Age at Menarche [yrs; mean (SD)]	12.4 (1.5)	12.8 (1.7)	12.5 (1.5)	12.8 (2.1)	12.6 (1.6)	12.8 (1.7)	12.6 (1.6)	12.8 (1.7)
Age at First Birth [yrs; mean (SD)]	23.9 (4.5)	23.7 (3.9)	23.9 (4.5)	23.7 (6.3)	24.3 (4.8)	24.1 (4.4)	23.8 (4.7)	23.5 (4.3)
Body Mass Index [mean; (SD)]	28.6 (5.9)	28.4 (6.1)	28.6 (5.8)	26.7 (6.3)	28.9 (5.8)	28.5 (5.9)	28.9 (6.0)	28.5 (6.1)
Age at Menopause [yrs; mean (SD)]	48.3 (5.0)	47.4 (6.0)	48.0 (5.3)	47.6 (6.0)	45.9 (5.6)	47.6 (6.0)	48.3 (5.4)	47.4 (6.3)
Benign Breast Disease (yes)	36%	22%	34%	23%	34%	21%	33%	22%
1° Relative with Breast Cancer (yes)	20%	14%	19%	15%	21%	13%	20%	14%

Table 2. Risk associated with exposure to Total Suspended Particulate Concentrations at Birth Address: Western New York Exposures and Breast Cancer Study (WEB Study).

TSP ug/m³	Premenopausal				Postmenopausal			
	Cases	Controls	Crude OR (95%CI)	Adjusted OR (95%CI)*	Cases	Controls	Crude OR (95%CI)	Adjusted OR (95%CI)*
	(n=164)	(n=283)			(n=357)	(n=521)		
<84	5	16	1.00	1.00	7	19	1.00	1.00
84-114	26	46	1.81 (0.59-5.51)	1.96 (0.64-3.01)	52	67	2.11 (0.82-5.39)	2.32 (0.89-6.10)
115-140	64	92	2.23 (0.78-6.39)	2.23 (0.77-6.44)	142	223	1.73 (0.71-4.21)	1.94 (0.77-4.86)
>140	69	129	1.71 (0.60-4.87)	1.78 (0.62-5.10)	156	215	1.97 (0.81-4.80)	2.42 (0.97-6.09)
P for Trend				0.38				0.01

* Adjusted for age, education, and parity.

Table 3. Risk associated with exposure to Total Suspended Particulate Concentrations at Menarche Address: Western New York Exposures and Breast Cancer Study (WEB Study).

TSP ug/m^3	Premenopausal				Postmenopausal			
	Cases	Controls	Crude OR (95%CI)	Adjusted OR (95%CI)*	Cases	Controls	Crude OR (95%CI)	Adjusted OR (95%CI)*
	(n=204)	(n=386)			(n=469)	(n=757)		
<84	32	66	1.00	1.00	14	28	1.00	1.00
84-114	53	99	1.10 (0.65-1.89)	0.98 (0.56-1.70)	81	120	1.35 (0.67-2.72)	1.36 (0.67-2.77)
115-140	62	81	1.58 (0.92-2.70)	1.25 (0.71-2.23)	171	298	1.15 (0.59-2.24)	1.20 (0.61-2.36)
>140	57	140	0.84 (0.50-1.42)	0.66 (0.38-1.16)	203	311	1.31 (0.67-2.54)	1.45 (0.74-2.87)
P for Trend				0.2114				0.1879

* Adjusted for age, education, and parity.

Table 4. Risk associated with exposure to Total Suspended Particulate Concentrations at First Birth Address: Western New York Exposures and Breast Cancer Study (WEB Study).

TSP ug/m³	Premenopausal				Postmenopausal			
	Cases	Controls	Crude OR (95%CI)	Adjusted OR (95%CI)*	Cases	Controls	Crude OR (95%CI)	Adjusted OR (95%CI)*
	(n=181)	(n=371)			(n=435)	(n=782)		
<84	147	294	1.00	1.00	54	102	1.00	1.00
84-114	19	30	1.27 (0.69-2.33)	1.06 (0.55-2.02)	89	150	1.12 (0.74-1.71)	1.30 (0.83-2.03)
115-140	5	19	0.53 (0.19-1.44)	0.41 (0.14-1.67)	142	260	1.03 (0.70-1.52)	1.28 (0.83-1.97)
>140	10	28	0.71 (0.34-1.51)	0.52 (0.22-1.20)	150	270	1.05 (0.71-1.24)	1.33 (0.87-2.06)
P for trend				0.0378				0.6064

* Adjusted for age, education, and parity.

APPENDIX IV

Environmental Tobacco Smoke Exposure in Early Life and the Risk of Breast Cancer

Matthew R. Bonner[1], Jing Nie[1], Daikwon Han[1], John E. Vena[2], Peter Rogerson[3], Paola Muti[1], Maurizio Trevisan[1], Stephen Edge[4], Jo L. Freudenheim[1]

[1] Department of Social and Preventive Medicine, School of Public Health and Health Professions, University at Buffalo, Buffalo, New York 14214.

[2] Department of Epidemiology and Biostatistics, Arnold School of Public Health, University of South Carolina, Columbia, South Carolina 29208.

[3] Department of Geography, University at Buffalo, Buffalo, New York 14214.

[4] Department of Surgery, Roswell Park Cancer Institute, Buffalo, NY 14263.

Corresponding author: Matthew R. Bonner, Occupational and Environmental Epidemiology Branch, National Cancer Institute, 6120 Executive Blvd. EPS 8121, MSC 7240, Bethesda, MD 20892-7240 Phone: (301) 402-7825, Fax: (301) 402-1819, E-mail address: bonnerm@mail.nih.gov

Abstract

Evidence is increasing that some early life exposures affect breast cancer risk. Exposure to environmental tobacco smoke (ETS) during childhood may be one such exposure. As part of the WEB Study (Western New York Exposures and Breast Cancer Study), we conducted a population-based, case-control study with 1,166 women aged 35-79 diagnosed with histologically confirmed, primary, incident breast cancer. Controls (n=2,105) were randomly selected from the Department of Motor Vehicles driver's license list (<age 65) and the Healthcare Financing Administration Medicare rolls (>age 65). Participants were queried regarding the number of smokers they lived with and the number of years they resided with these smokers. Person-years of ETS exposure was computed. Unconditional logistic regression adjusting for potential confounders was used to calculate odds ratios (OR) and 95% confidence intervals (95% CI). Exposure to ETS before the age of 21 may be weakly associated with an increase in breast cancer risk for premenopausal women (OR = 1.46, 95% CI: 0.99- 2.15) and postmenopausal (OR = 1.22; 95% CI: 0.94-1.58) women. Although these estimates may suggest a weak association between early life ETS exposure and breast cancer, we cannot exclude the possibility that ETS exposure is unrelated to risk.

Introduction

Increasingly, there is interest that exposures in early life may be related to breast cancer risk [1-3] because of evidence that the breast may be more vulnerable to carcinogenic insults during this period of tissue proliferation and initial differentiation. Pregnancy and lactation result in further differentiation after which it seems that breast tissue is more resistant to carcinogenic insults.[4-6] It has been hypothesized that exposures before a woman gives birth for the first time may be particularly important in relation to disease etiology.[7, 8] Environmental tobacco smoke is one such exposure that may affect the risk of developing breast cancer.

Exposure to environmental tobacco smoke is relatively common among U.S. children where an estimated 43% reside with at least one smoker.[9] Tobacco smoke consists of numerous compounds that are carcinogenic to several organ sites including the lung [10-12], bladder [13], and pancreas.[14] Among these compounds are polycyclic aromatic hydrocarbons (PAHs) and aromatic amines. PAHs are known skin and mammary carcinogens in rodent models [15-17] and accumulate in adipose tissue including the breast.[18, 19] In addition, aromatic amines have been shown to be mammary carcinogens in rodents. [20] The effect of tobacco smoke on breast cancer risk, however, is not clear. McMahon [21] hypothesized that tobacco smoke may reduce the risk of breast cancer because of evidence that cigarette smoke has antiestrogenic effects. Conversely, Hiatt and Fireman [22] reasoned that smoking could increase breast cancer risk because mutagens from cigarette smoke concentrate in the breast fluid of nonlactating women. Cigarette smoking is also associated with bladder [13] and pancreatic [14] cancers, all sites without direct contact between smoke and the organ's epithelium. Despite conflicting

hypotheses about the effect of tobacco smoke on breast cancer risk, an association between cigarette smoking and breast cancer has yet to be clearly demonstrated.[23] Further, there is limited evidence that environmental tobacco smoke (ETS) affects breast cancer risk.[24, 25] With regards to early life exposure, there have been a few studies of ETS and risk; ETS has been found to be associated with increased risk of breast cancer in some [26], but not in all studies.[27-29]

In this study, we explored exposure to ETS in early life in relation to the risk of subsequent breast cancer. Specifically, we hypothesized that residing with one or more household smokers in early life (up to age 21) would increase the risk of breast cancer compared with women who did not reside with household smokers.

MATERIAL AND METHODS

The Western New York Exposures and Breast Cancer Study (WEB Study) is a population-based, case-control study in Western New York. Cases (n=1,166) included women aged 35-79 years diagnosed with histologically-confirmed, primary, incident breast cancer currently residing in Erie or Niagara Counties in Western New York. Nurse case finders visited the pathology departments at regular intervals to identify cases. After identification, the case's physician was contracted to verify the diagnosis of breast cancer and to obtain permission to contact the case. Cases were then contacted and asked to participate in the study. Controls were also current residents of Erie and Niagara counties randomly selected from the New York State Department of Motor Vehicles driver's license list (aged 65 and less) and the Centers for Medicare and Medicaid Services rolls (over 65 years). Controls ($n = 2,105$) were frequency matched by age and race. A total of 1,638 cases and 3,396 controls met our inclusion criteria of between 35-

79 years of age, current resident of Erie or Niagara County, no previous cancer diagnosis other than non-melanoma skin cancer and an ability to speak English. The response rates were 71% (1,166/1,638) and 62% (2,105/3,396) for cases and controls, respectively. All participants provided informed consent; the protocol was approved by the University at Buffalo School of Medicine and Biomedical Sciences' and participating hospitals' Institutional Review Boards

Extensive in-person interviews and self-administered questionnaires were used to ascertain medical history, diet, lifetime alcohol consumption, residential history, occupational history, and smoking history. We evaluated exposure to ETS with two methods. First, questions about exposure to environmental tobacco smoke were asked for seven age periods: 1) <21 years, 2) 21-30, 3) 31-40, 4) 41-50, 5) 51-60, 6) 61-70, 7) >70. The number of people living with the participant who smoked cigarettes, cigars, or pipes during the specified time period was ascertained. In addition, participants were also asked for the number of years that they resided with these smokers. These two questions were used to compute person-years of ETS exposure for participants for each time period. For this study, we only considered exposure to ETS before 21 years of age. Exposure to ETS was categorized into three groups: 1) no ETS exposure (0 person-years), 2) >0 to ≤20 person-years of ETS exposure, and 3) >20 person-years of ETS exposure. The cut point of 20 person-years of ETS exposure was derived from the median in the exposed controls.

The second evaluation of ETS exposure was part of the residential history assessment. Participants listed each residence for their entire life with corresponding information on the number of other people who resided at that residence and the number

of those residents who smoked cigarettes. The analysis was restricted to those with complete household smoking information at both birth and menarche ($n = 334$ for cases and 609 for controls). For exposure at the time of first birth, the analysis was further restricted to those with residential information for all three time periods. Household smoking was categorized into a binary variable denoting either the presence or absence of household smokers.

Unconditional logistic regression [30] was used to calculate odds ratios (OR) and 95 % confidence intervals (CI), adjusting for age, race, education, age at first birth, age at menarche, parity, previous benign breast disease, family history of breast cancer in a first degree relative, body mass index (weight (kg)/height (m)^2), pack-years of smoking, total lifetime alcohol consumption, and age at menopause for postmenopausal women only. All models were stratified by menopausal status. A reduced model including age previous benign breast disease, and pack-years of smoking was determined by removing covariates that did not alter the OR by more than 10%. Additional analyses were conducted excluding former and current smokers to prevent a history of active smoking from confounding any potential association between ETS and the risk of breast cancer. *P* for trend statistics was determined by the *p*-value for the coefficient of the continuous exposure variable, while adjusting for covariates.

RESULTS

Demographic characteristics of the study participants by menopausal status are shown in table 1. Exposure to ETS before the age of 21 was associated with an increase in the risk of breast cancer for both premenopausal women (reduced model OR = 1.46, 95% CI: 0.99, 2.15) and postmenopausal women (reduced model OR = 1.22, 95% CI:

0.94-1.58) (table 2); although, confidence intervals include the null for both groups. When the analysis was restricted to the sub-group of never smokers, similar results were obtained, although confidence intervals were wider because of the decrease in sample size (data not shown).

Associations between the presence of household smokers in the participant's residence at the time of their birth, menarche, and first birth and breast cancer among never smokers are shown in table 3. There was some tendency for premenopausal women with breast cancer to reside with one or more household smokers at their birth address more often than controls (reduced model OR = 1.40, 95% CI: 0.83-2.38), while in postmenopausal women, the presence of household smokers was not associated with breast cancer (reduced model OR = 1.02, 95% CI: 0.68-1.55). Associations between the presence of household smokers at the time of menarche and breast cancer were similar.

The presence of household smokers at the time of a women's first birth was not associated with breast cancer in premenopausal women (reduced model OR = 1.24, 95% CI: 0.70-2.21). For postmenopausal women, however, exposure to household smoke at the time of first birth was suggestive, if anything, of a reduction in risk (reduced OR = 0.71, 95% CI: 0.46-1.09). We attempted to examine each time period while adjusting for the other two time periods to investigate whether one time period in particular was associated with an increased odds ratio. However, household smoking status at each of the time periods was highly correlated and the results were not interpretable.

DISCUSSION

Overall, this study provides little evidence that exposure to ETS in early life is associated with an increase in the risk of breast cancer. In the few studies that have

examined early life exposure to ETS, the results have been mixed. In one study, Sandler et al. [28] found no increase in the risk of breast cancer in women exposed to either maternal or paternal household smoking before participants attained 10 years of age. In another study, Smith et al. [29] assessed exposure to ETS up to age 16 and found women exposed to ETS only in childhood had an OR of 1.98 (95% CI: 0.35-11.36) compared with those never exposed. In addition, women exposed in childhood to 201-400 cigarette-years were observed to have an OR of 2.09 (95% CI: 1.05-4.16). However, the OR was 1.51 (95% CI: 0.72-3.20) in women exposed in childhood to >400 cigarette-years. Smith et al. concluded that there was no association between ETS exposure in childhood and breast cancer. Lash et al. [26] examined women who were exposed to ETS before the age of 12 and found an OR of 4.5 (95% CI: 1.2-16). For women exposed to ETS alone, the OR was 7.5 (95% CI: 1.6-36). However, these results were not replicated in a more recent case-control study where exposure to ETS before the age of 13 was not associated with an increase in the risk of breast cancer (OR: 1.1, 95% CI: 0.4-3.0).[27]

To our knowledge, this study is the first to specifically examine ETS exposure at birth, menarche, and first birth. However, interpretation of these results was difficult because of the modest risk estimates and wide confidence intervals. The estimated ORs suggest there may be a weak association with exposure to household smokers in early life, particularly around the time of birth and menarche.

Tobacco smoke contains potent carcinogens and there is evidence that these are deposited in breast tissue.[18, 19] Consequently, it is biologically plausible that ETS could be a risk factor for breast cancer. Furthermore, the timing of exposure may be crucial in defining the role of ETS and active smoking in the etiology of breast cancer.[26] In

addition to PAHs and other carcinogenic compounds, cigarette smoke also contains carbon monoxide, which is an inhibitor of the cytochrome P-450s. It is possible that by inhibiting the cytochrome P-450s, carbon monoxide prevents PAHs from being metabolized to their ultimate carcinogen moiety.[19] If this is occurring, then PAHs in cigarette smoke would not contribute to carcinogenesis and the use of smoking as a surrogate for PAH exposure would be misleading. Conversely, cigarette smoke has been hypothesized to be anti-estrogenic and therefore may reduce the risk of breast cancer, although our results do not support this hypothesis. It may be that genetically heterogeneous study populations have obscured the net effect that cigarette smoke may have on breast cancer risk.[31, 32]

There are several limitations of this study that should be considered when interpreting the results. Among these is recall bias. While such a bias is possible, it would seem less likely, given the request for information pertained to childhood experiences and that there is no well known hypothesis linking ETS exposure in early life and breast cancer risk. Misclassification of exposure is likely given that ETS was crudely measured and did not take into account other sources of ETS. Further misclassification of ETS exposure could have occurred because the some smokers may not have smoked in the presence of that participant. In particular, we could not distinguish smokers who restricted their smoking activities around the participant, thereby decreasing exposure, from those who did not. In addition, we assumed that early life exposure to ETS would predominantly occur in the household. This is particularly likely for the time period between birth and menarche. Regardless, this measure is not quantitative and the potential for non-differential misclassification of exposure exists.

In addition, the possibility of selection bias cannot be ruled out. Comparisons between respondents and non-respondents indicated that smokers were less likely to participate in this study. Since smokers are more likely to have parents who smoked [33], a selection bias may have altered the distribution of ETS exposure in the controls from that of the source population from which the cases arose resulting in magnified risk estimates.

In summary, our study examined exposure to household tobacco smoke up to the 21 years of age and at the time of birth, menarche, and first birth in relation to the development of breast cancer. We had hypothesized that the chemical carcinogens present in tobacco smoke such as PAHs would affect breast cancer risk and that exposure to tobacco smoke in early life would have particular importance. Although these estimates may suggest a weak association between early life ETS exposure and breast cancer, we cannot exclude the possibility that ETS exposure is unrelated to risk. The recent trends toward limiting ETS exposure particularly for children remains appropriate, given our knowledge of other effects of ETS on health [34] and the relatively high prevalence of ETS exposure in the U.S. population.[9]

REFERENCES

1. Potischman N, Troisi R. In-utero and early life exposures in relation to risk of breast cancer. *Cancer Causes Control* 1999;10(6):561-73.

2. Trichopoulos D. Is breast cancer initiated in utero? *Epidemiology* 1990;1(2):95-6.

3. Trichopoulos D. Hypothesis: does breast cancer originate in utero? *Lancet* 1990;335(8695):939-40.

4. Russo J, Reina D, Frederick J, Russo IH. Expression of phenotypical changes by human breast epithelial cells treated with carcinogens in vitro. *Cancer Res* 1988;48(10):2837-57.

5. Russo J, Calaf G, Sohi N, Tahin Q, Zhang PL, Alvarado ME, et al. Critical steps in breast carcinogenesis. *Ann N Y Acad Sci* 1993;698:1-20.

6. Russo J, Hu YF, Yang X, Russo IH. Developmental, cellular, and molecular basis of human breast cancer. *J Natl Cancer Inst Monogr* 2000(27):17-37.

7. Berkey CS, Frazier AL, Gardner JD, Colditz GA. Adolescence and breast carcinoma risk. *Cancer* 1999;85(11):2400-9.

8. Colditz GA, Frazier AL. Models of breast cancer show that risk is set by events of early life: prevention efforts must shift focus. *Cancer Epidemiol Biomarkers Prev* 1995;4(5):567-71.

9. Pirkle JL, Flegal KM, Bernert JT, Brody DJ, Etzel RA, Maurer KR. Exposure of the US population to environmental tobacco smoke: the Third National Health and Nutrition Examination Survey, 1988 to 1991. *Jama* 1996;275(16):1233-40.

10. Lubin JH. Estimating lung cancer risk with exposure to environmental tobacco smoke. *Environ Health Perspect* 1999;107 Suppl 6:879-83.

11. Bennett WP, Alavanja MC, Blomeke B, Vahakangas KH, Castren K, Welsh JA, et al. Environmental tobacco smoke, genetic susceptibility, and risk of lung cancer in never-smoking women. *J Natl Cancer Inst* 1999;91(23):2009-14.

12. Jee SH, Ohrr H, Kim IS. Effects of husbands' smoking on the incidence of lung cancer in Korean women. *Int J Epidemiol* 1999;28(5):824-8.

13. Brennan P, Bogillot O, Cordier S, Greiser E, Schill W, Vineis P, et al. Cigarette smoking and bladder cancer in men: a pooled analysis of 11 case-control studies. *Int J Cancer* 2000;86(2):289-94.

14. Kuper H, Boffetta P, Adami HO. Tobacco use and cancer causation: association by tumour type. *J Intern Med* 2002;252(3):206-24.

15. Bostrom CE, Gerde P, Hanberg A, Jernstrom B, Johansson C, Kyrklund T, et al. Cancer risk assessment, indicators, and guidelines for polycyclic aromatic hydrocarbons in the ambient air. *Environ Health Perspect* 2002;110 Suppl 3:451-88.

16. Phillips DH. Fifty years of benzo(a)pyrene. *Nature* 1983;303(5917):468-72.

17. Rubin H. Synergistic mechanisms in carcinogenesis by polycyclic aromatic hydrocarbons and by tobacco smoke: a bio-historical perspective with updates. *Carcinogenesis* 2001;22(12):1903-30.

18. Obana H, Hori S, Kashimoto T, Kunita N. Polycyclic aromatic hydrocarbons in human fat and liver. *Bull Environ Contam Toxicol* 1981;27(1):23-7.

19. Morris JJ, Seifter E. The role of aromatic hydrocarbons in the genesis of breast cancer. *Med Hypotheses* 1992;38(3):177-84.

20. Dunnick JK, Elwell MR, Huff J, Barrett JC. Chemically induced mammary gland cancer in the National Toxicology Program's carcinogenesis bioassay. *Carcinogenesis* 1995;16(2):173-9.

21. MacMahon B, Trichopoulos D, Cole P, Brown J. Cigarette smoking and urinary estrogens. *N Engl J Med* 1982;307(17):1062-5.

22. Hiatt RA, Fireman BH. Smoking, menopause, and breast cancer. *J Natl Cancer Inst* 1986;76(5):833-8.

23. Palmer JR, Rosenberg L. Cigarette smoking and the risk of breast cancer. *Epidemiol Rev* 1993;15(1):145-56.

24. Egan KM, Stampfer MJ, Hunter D, Hankinson S, Rosner BA, Holmes M, et al. Active and passive smoking in breast cancer: prospective results from the Nurses' Health Study. *Epidemiology* 2002;13(2):138-45.

25. Morabia A, Bernstein M, Heritier S, Khatchatrian N. Relation of breast cancer with passive and active exposure to tobacco smoke. *Am J Epidemiol* 1996;143(9):918-28.

26. Lash TL, Aschengrau A. Active and passive cigarette smoking and the occurrence of breast cancer. *Am J Epidemiol* 1999;149(1):5-12.

27. Morabia A, Lash TL. Breast cancer, passive and active cigarette smoking and N-acetyltransferase 2 genotype. *Pharmacogenetics* 2002;12(1):85-8.

28. Sandler DP, Everson RB, Wilcox AJ, Browder JP. Cancer risk in adulthood from early life exposure to parents' smoking. *Am J Public Health* 1985;75(5):487-92.

29. Smith SJ, Deacon JM, Chilvers CE. Alcohol, smoking, passive smoking and caffeine in relation to breast cancer risk in young women. UK National Case-Control Study Group. *Br J Cancer* 1994;70(1):112-9.

30. Hosmer DW and Lemeshow S. Applied Logistic Regression. New York: Wiley, 2000.

31. Ambrosone CB, Freudenheim JL, Graham S, Marshall JR, Vena JE, Brasure JR, et al. Cigarette smoking, N-acetyltransferase 2 genetic polymorphisms, and breast cancer risk. *Jama* 1996;276(18):1494-501.

32. Wartenberg D, Calle EE, Thun MJ, Heath CW, Jr., Lally C, Woodruff T. Passive smoking exposure and female breast cancer mortality. *J Natl Cancer Inst* 2000;92(20):1666-73.

33. Bauman KE, Foshee VA, Linzer MA, Koch GG. Effect of parental smoking classification on the association between parental and adolescent smoking. *Addict Behav* 1990;15(5):413-22.

34. American Academy of Pediatrics, Environmental tobacco smoke: a hazard to children. American Academy of Pediatrics Committee on Environmental Health, 1997.

TABLE 1. Descriptive Characteristics of Study Participants: Western New York Exposures and Breast Cancer Study (WEB Study) (1996-2001).

	Premenopausal		Postmenopausal	
	Cases	Controls	Cases	Controls
	(n=325)	(n=610)	(n=841)	(n=1495)
Age				
35-45	176 (54%)	384 (63%)	7 (1%)	29 (2%)
46-55	149 (46%)	224 (37%)	175 (21%)	325 (22%)
56-65	-	2 (1%)	325 (39%)	403 (27%)
66-75	-	-	262 (31%)	630 (42%)
76+	-	-	72 (9%)	108 (7%)
Education				
<High school	2 (1%)	-	22 (3%)	38 (3%)
High school	105 (32%)	176 (32%)	388 (46%)	780 (52%)
>High school	218 (67%)	434 (71%)	431 (51%)	677 (45%)
Age at Menarche				
<12	80 (25%)	134 (22%)	199 (24%)	327 (22%)
12-13	212 (65%)	414 (68%)	548 (65%)	966 (65%)
14+	33 (10%)	62 (10%)	94 (11%)	202 (14%)
Age at Menopause				
<45	-	-	157 (19%)	389 (26%)
45-49	-	-	222 (26%)	374 (25%)
50-54	-	-	373 (44%)	582 (39%)
55+	-	-	89 (11%)	150 (10%)

TABLE 1, continued. Descriptive Characteristics of Study Participants: Western New York Exposures and Breast Cancer Study (WEB Study) (1996-2001).

	Premenopausal		Postmenopausal	
	Cases (*n*=325)	Controls (*n*=610)	Cases (*n*=841)	Controls (*n*=1495)
Age at First Birth				
Never	58 (18%)	99 (16%)	148 (18%)	155 (10%)
13-19	45 (14%)	38 (6%)	91 (11%)	203 (14%)
20-21	29 (9%)	62 (10%)	150 (18%)	259 (17%)
22-25	73 (22%)	157 (26%)	271 (32%)	512 (34%)
26-39	120 (37%)	254 (42%)	181 (22%)	366 (24%)
Parity				
0	58 (18%)	99 (16%)	148 (18%)	155 (10%)
1-2	177 (54%)	323 (53%)	293 (35%)	452 (30%)
3+	90 (28%)	188 (31%)	400 (48%)	888 (59%)
Body Mass Index				
<25	144 (44%)	266 (44%)	243 (29%)	448 (30%)
25-29	93 (29%)	168 (28%)	277 (33%)	567 (38%)
30+	88 (27%)	176 (29%)	321 (38%)	480 (32%)
Benign Breast Disease (yes)	120 (37%)	130 (21%)	278 (33%)	327 (22%)
Relative with Breast Cancer (yes)	61 (21%)	56 (10%)	160 (20%)	196 (14%)
Smoking Status				
Never	149 (46%)	326 (54%)	376 (45%)	686 (46%)
Former	123 (38%)	182 (30%)	365 (44%)	584 (39%)
Current	53 (16%)	101 (17%)	99 (12%)	224 (15%)

TABLE 2. Risk of Breast Cancer Associated with Exposure to Environmental Tobacco Smoke before the age 21: Western New York Exposures and Breast Cancer Study (WEB Study) (1996-2001).

			Premenopausal						Postmenopausal				
			Reduced Model		Full Model				Reduced Model		Full Model	
	Cases	Controls	OR*†	95% CI*†	OR‡	95%CI‡	Cases	Controls	OR	95%CI	OR	95%CI
	(n=325)	(n=609)					(n=841)	(n=1491)				
ETS* (person-years)												
0	60	153	1.00	Ref.	1.00	Ref.	173	360	1.00	Ref.	1.00	Ref.
>0-≤20	145	259	1.37	0.95, 1.98	1.32	0.91, 1.93	458	774	1.21	0.98, 1.51	1.23	0.99, 1.54
>20	120	197	1.46	0.99, 2.15	1.44	0.97, 2.14	210	357	1.22	0.94, 1.58	1.22	0.94, 1.59
P for trend				0.41		0.47				0.28		0.21

*ETS, environmental tobacco smoke; OR, odds ratio; CI, confidence interval.

† Adjusted for age, pack-years of smoking, and previous benign breast disease

‡ Adjusted for age, education, race, previous benign breast disease, parity, age at menarche, BMI, age at first birth, relative with breast cancer, pack-years of smoking, total alcohol consumption and age at menopause for postmenopausal women only.

TABLE 3. Risk of Breast Cancer Associated with Exposure to Environmental Tobacco Smoke Exposure at the Time of Birth Menarche and First Birth; among never smokers: Western New York Exposures and Breast Cancer Study (WEB Study) (1996-2001).

| | Premenopausal | | | | | | Postmenopausal | | | | | |
| | Cases | Controls | Reduced Model | | Full Model | | Cases | Controls | Reduced Model | | Full Model | |
ETS* exposure	$(n=106)$	$(n=238)$	OR*†	95% CI*†	OR‡	95% CI‡	$(n=228)$	$(n=371)$	OR	95% CI	OR	95% CI
Birth												
No	27	84	1.00	Ref.	1.00	Ref.	52	89	1.00	Ref.	1.00	Ref.
Yes	79	154	1.40	0.83, 2.38	1.34	0.77, 2.32	176	282	1.02	0.68, 1.55	1.07	0.72, 1.59
Menarche												
No	29	92	1.00	Ref.	1.00	Ref.	50	89	1.00	Ref.	1.00	Ref.
Yes	77	146	1.54	0.92, 2.58	1.49	0.87, 2.57	178	282	1.05	0.69, 1.60	1.15	0.77, 1.71
First Birth§												
No	52	140	1.00	Ref.	1.00	Ref.	66	108	1.00	Ref.	1.00	Ref.
Yes	31	56	1.24	0.70, 2.21	1.11	0.59, 2.11	93	188	0.71	0.46, 1.09	0.80	0.54, 1.20

* ETS, environmental tobacco smoke; OR, odds ratio; CI, confidence interval.

†Adjusted for age and previous benign breast disease; ‡Adjusted for age, race, education, previous benign breast disease, parity, age at menarche, BMI, age at first birth, family history of breast cancer, total alcohol consumption, and age at menopause for postmenopausal women only.

§ Restricted to cases and controls with known addresses at the time of birth, menarche and first birth.

APPENDIX V

Positional Accuracy of Geocoding in Epidemiologic Research. Matthew R. Bonner[1], Daikwon Han[2], Jing Nie[1] Jo L. Freudenheim[1], Peter Rogerson[2],John E. Vena[1]

Geographic Information Systems (GIS) offer powerful techniques for epidemiologists. Geocoding is an important step in the use of GIS in epidemiologic research and the validity of any epidemiologic study using this methodology depends, in part, on the positional accuracy of the geocoding process. We conducted a study comparing the validity of positions geocoded with a commercially available program to positions determined by receivers for the Global Positioning System (GPS) satellites.

Methods:

Addresses (n=200) were randomly selected from a recently completed case-control study in Western New York. These addresses were geocoded using ArcView 3.2 on the GDT Dynamap/2000 U.S. Street database. Latitude and longitude of these same addresses were measured with a GPS receiver, and distance between these two points was calculated for all addresses.

Results:

The distance between the geocoded point and the GPS point was within 100m for the majority of the all subject addresses (79%) with only a small proportion (3%) having a distance greater than 800m. The overall median distance between GPS points and geocoded points was 38m (90% CI 33.67-45.90). Distances were not different for cases and controls. Urban addresses (32m; 90% CI 28.32-36.81) were slightly more accurate compared to the non-urban addresses (52m; 90% CI 43.51-61.06).

Conclusions: .

Overall, this study indicates that the suitability of geocoding for epidemiologic research depends on the level of spatial resolution required to assess exposure. While sources of error in positional accuracy for geocoded addresses exist, geocoding of addresses is largely very accurate.

Household smoke exposure in early life and breast cancer in Western New York.
Bonner MR, Nie J, Han D, Vito D,Vena JE, Rogerson P, Muti P, Trevisan M,
Freudenheim JL. American Association for Cancer Research, Toronto, Canada,
April, 2003.

Exposure to tobacco smoke in early life may be more relevant for breast cancer than exposure in adult life. Numerous epidemiologic studies of adult smoking exposure have been equivocal. Relatively few investigations, however, have examined tobacco smoke exposure in early life when breast epithelium may be more sensitive to carcinogens. In this study, we hypothesized that household tobacco smoke exposure during critical time periods of breast development (birth, menarche, and first birth) may be associated with the occurrence of breast cancer. As part of the Center for Preventive Medicine, we used a case-control study design with 1,170 cases of primary, histologically confirmed, incident breast cancer and 2,116 population-based controls. Exposure to household smokers at birth, at menarche and at first birth was assessed with a self-administered residential history questionnaire. Each subject indicated all previous residences as well as the number of other people residing at that address and the number of those household residents who smoked. We categorized the number of household smokers into none, one, and two or more household smokers. Logistic regression was used to estimate odds ratios (OR) and 95% confidence intervals (95% CI) with no household smokers as the referent category. Multivariate logistic models were adjusted for age at interview, years of education, previous benign breast disease, age at menarche, parity, body mass index, total lifetime alcohol consumption, and relative with breast cancer. We found that women who at birth resided with 1 or more household smokers, were more likely to develop breast cancer compared to those residing with no household smokers (adjusted OR = 1.36, 95% CI = 1.08-1.70). A similar association was also observed for women who at menarche were exposed to household smokers (adjusted OR = 1.43, 95% CI =1.15-1.77). Exposure to household smokers at the time of first birth was more weakly associated with breast cancer (adjusted OR = 1.19, 95% CI = 0.96-1.47). In a logistic model simultaneously adjusting for household smoke exposure at all three time periods, only exposure to household smokers at the time of menarche remained above unity (adjusted OR = 1.77, 95% CI = 1.00-3.15). However, exposure to household smoke in these time periods tended to be correlated. These results suggest that household smoke exposure in early life may be associated with an increase in the likelihood of breast cancer and it may be that exposure at the time of menarche is more important than exposure at other time periods.

Clustering of Lifetime Residence and Breast Cancer Risk in Western New York

*D Han, MR Bonner, J Nie, PA Rogerson, JE Vena, P Muti, M Trevisan, JL Freudenheim. University at Buffalo, Buffalo, NY 14214.

In order to investigate the role of environmental exposures on breast cancer, we examined breast cancer risk associated with lifetime residential history using GIS-based exploratory spatial analyses. Data on residential history and risk factors were collected as part of a population-based case control study of incident, primary, histologically-confirmed breast cancer in western New York. Controls were frequency matched to cases on age and county of residence. Relative risk surfaces of cases and controls were identified to depict elevated areas of breast cancer risk using kernel smoothing methods. The ratio of cases to controls was first obtained based on location of their residence for each participant at the time of birth, menarche, first birth, and 10 and 20 years before interview, then adjusted for established breast cancer risk factors using a generalized additive model. Cumulative risk surfaces were constructed by using case-control densities from each temporal group. These surfaces were compared between residences for pre-menopausal and post-menopausal women. We found a general tendency of spatial clustering of lifetime residence, and we observed strong evidence of clustering of lifetime residence for pre-menopausal women relative to that for post-menopausal women. We were able to pinpoint geographic areas with higher cumulative densities, but also to identify the role of early exposures through exploratory spatial analyses. Our findings suggest that there may be identifiable etiological processes on exposure and breast cancer risk, especially for pre-menopausal women, and that early exposures may be of particular importance.

Total Suspended Particulate Exposure in Early Life and Breast Cancer. *MR Bonner, D Han, J Nie, JE Vena, P Rogerson, P Muti, M Trevisan, D Vito, JL Freudenheim. University at Buffalo, Buffalo, NY 14214.

Polycyclic aromatic hydrocarbons (PAHs) are ubiquitous environmental pollutants present in air pollution and largely associated with particulate matter. PAHs may be estrogenic and could contribute to breast cancer etiology. Further, early life exposures may be significant in the development of this disease. We examined total suspended particulate (TSP) exposure (as a proxy for PAH exposure) in early life in relation to the risk of breast cancer. We conducted a population-based case-control study with 1,170 cases of primary, histologically confirmed, incident breast cancer and 2,116 randomly selected controls. TSP concentrations measured by air monitoring samplers from 1958-1991 in Erie and Niagara counties were used to estimate TSP exposure. Average TSP concentrations were computed for each decade and inverse distance squared weighting interpolation was used to estimate TSP concentrations for each subject's residence at birth, menarche, and first birth. Logistic regression was used to estimate odds ratios (OR) and 95% confidence intervals (95% CI), adjusting for potential confounders. No association in risk was observed in premenopausal women for exposure to TSP. In postmenopausal women, the continuous adjusted OR was 1.21 (95% CI 1.05-1.40) for every 30 mg/m^3 increase in exposure to TSP at the birth residence. In this group, risk associated with exposure of over 135 mg/m^3 of TSP exposure at the time of birth compared with postmenopausal women with <81 mg/m^3, was 2.59 (95% CI 0.96-7.03). These results suggest that high levels of exposure in early life to TSP may be associated with an increase in the risk of postmenopausal breast cancer.

Daikwon Han, "Geographical Epidemiology of Breast Cancer in Western New York: Migration and Disease Clustering," Annual Meeting of the Association of American Geographers, Los Angeles, CA. 2002.

Migration has a significant effect on geographic variations of disease and health outcomes. The complex process of human movement is one of the complicating factors in explaining the causal relationships between disease and environment, but also an important determinant of human health due to the exposure to disease through movement. This study explores the migration effects on disease clustering to assess; 1) the importance of residential locations to the risk of breast cancer, 2) the statistical significance of clustering with migration effects. To identify the reasons for geographic variations of disease, the study presents hypotheses associated with migration and disease risks. Exploratory analyses in a GIS environment are used to detect the spatial-temporal patterns of residential locations and clustering of case-controls in Western New York. The overall effects of migration on disease clustering are identified by comparing the lifetime residential history of case-controls, after controlling for the known risk factors such as age and history of breast cancer. The investigation on the role of migration on disease clustering processes provide explanations on the consequences of in- and out-movement of people diagnosed with disease on the risk of disease as well as on the spatial variations of disease. Once significant clusters are identified, further work is required to investigate the relationships between residential changes and environmental exposures in explaining unknown etiology of breast cancer.

Residential Proximity at Birth to Industrial Sites and Subsequent Risk of Breast Cancer, MR Bonner, D Han, J Nie, JL Freudenheim, JE Vena, American College of Epidemiology Annual Meeting, September, 2002, Albuquerque, NM.

Purpose: To investigate the relationship between residential proximity at birth to industrial sites contracted by the Atomic Energy Commission (AEC) to process radioactive material and the subsequent development of breast cancer (BC) in pre and post menopausal women.

Methods: We used a completed case-control study (n=3,335) and restricted subjects to lifetime residents of Western New York (n=1,181). Subjects were further restricted to those born in 1940 and later because the first industrial sites began operating under the AEC contract in 1940. A total of 266 primary incident breast cancer cases and 411 controls frequency matched by age were included in this analysis. Exposure was assessed as distance (in miles) of residence at birth to the 13 industrial sites. The closest site was then selected for each subject as a surrogate for environmental exposure. The distance to the closest site was categorized into quartiles based on the distribution in the controls. Odds ratios (OR) and 95% confidence intervals (95% CI) were used to estimate the association between residential proximity and subsequent BC. The ORs were adjusted for age, education, age at menarche, parity, and age at first birth.

Results: We observed an adjusted OR of 3.8 (95% CI 1.9-7.7) for premenopausal women residing less than 2.45 miles from the closest site when compared to women residing greater than 8 miles from the closest industrial site. No such associations were observed in post menopausal women.

Conclusion: These preliminary findings suggest that relatively close residential proximity to industrial sites involved in uranium processing may increase the risk of premenopausal BC. However, it is unclear whether this association can be attributed to the environmental contamination with radioactive material, or some other environmental contaminate also produced at these industrial sites.

Daikwon Han, Jing Nie, Matthew Bonner, Dominica Vito, Jo Freudenheim,
"Environmental Exposures Associated with Lifetime Residential History: A GIS-based
Clustering Analysis of Breast Cancer," Annual Meeting of the Society for Epidemiologic

There is increasing evidence that early exposures may be related to risk of breast cancer. We were interested in whether there was clustering of breast cancer based on their residence in early life and identified spatio-temporal clustering of cases and controls at critical time periods, residential locations at birth, at menarche, and at the women's first birth. Data used here were part of the Center for Preventive Medicine case control study of incident, pathologically confirmed breast cancer (1996-2001) in Erie and Niagara counties. Controls were frequency matched on age and county of residence; controls less than 65 were randomly selected from the New York State Department of Motor Vehicles list and those greater than 65 from the Health Care Finance Administration list. All cases and controls provided lifetime residential histories. The spatial k-function method was used to calculate the distance between each residence within a certain search radius and to compare observed with expected patterns over pre-specified distances. We found a general tendency of spatial clustering for cases for these time periods, especially at small geographic scales, compared with the simulated theoretical distribution of expected patterns. The evidence for clustered residence at birth and at menarche was stronger than that for first birth. This study provides additional evidence that early environmental exposures may be related to breast cancer risk.

"Exploratory Spatial Analyses of Lifetime Breast Cancer Risk and Residence History" Daikwon Han, Jo L. Freudenheim, Peter A. Rogerson, Matthew R. Bonner, Jing Nie. Annual Meeting of the Association of American Geographers, New Orleans, LA. 2003.

This research investigates lifetime breast cancer risk associated with residential history based upon epidemiologic methods and exploratory spatial analyses. Data were drawn from a case control study of breast cancer in western New York and provided information on lifetime residential history and risk factors for 1170 breast cancer cases and 2116 controls. Epidemiologic methods were utilized to identify relationships between breast cancer risk and residence history. The ratio of cases to controls was obtained based on residential location and these ratios were adjusted for established risk factors, including age, education, and history of benign breast disease. Density surfaces of cases and controls were created to identify elevated areas of breast cancer risk using kernel smoothing methods, and these were repeated for six temporal groups; residences at birth, at menarche, at women's first birth, 20 years prior to diagnosis, 10 years prior to diagnosis, and current addresses. Lifetime risk surfaces were constructed and visualized by using case-control densities from each temporal group. These surfaces were further analyzed using weights dependent upon length of residence.

"Residential Proximity to Chemical or Primary Metal Industry and the Risk of Breast Cancer in Western New York" *J. Nie, Bonner M, Han D, LaFalce J, Vena JE, Freudenheim JL Presented at the Annual Meeting of the Society for Epidemiologic Research, Atlanta, GA, June 2003.

Women living in urban environments are at greater risk of breast cancer than those in rural settings; this difference is not well understood. In this study, we examined residential proximity to chemical or primary metal industry in relation to breast cancer risk. Women, age 35-79 with incident, primary, histologically confirmed breast cancer living in Erie or Niagara counties were invited to participate; and controls were population based, frequency matched to cases on age and race. Self-reported lifetime residential histories were collected. 863 cases and 1579 controls with complete residential addresses for the periods 10 and 20 years prior to interview were included studying these analyses. Industrial directories for New York State for 1978 and 1988, were used identify chemical and primary metal factories operating in this region. The chemical facility in our study includes Standard Industrial Classification (SIC) groups 28(Chemicals and allied products), 29(Petroleum refining and related industries), and 30(Rubber and miscellaneous plastics products); and primary metal facility (SIC 33). Quartiles were created to categorize the distance from residential address to the closest industrial site; women living within 0.25 mile of a facility were put in a separate category. We used logistic regression to calculate the odds ratios and 95% confidence intervals, adjusting for potential confounding factors. For both time periods and for both pre- and postmenopausal women, there was no evidence that living close to chemical or primary metal facility 10 and 20 years ago was associated with increased breast cancer risk.